

FABRÍCIO MARTINS LOPES

## **UM MODELO PERCEPTIVO DE LIMIAÇÃO DE IMAGENS DIGITAIS**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Informática pelo Curso de Pós-Graduação em Informática, do Setor de Ciências Exatas da Universidade Federal do Paraná, em convênio com o Departamento de Informática da Universidade Estadual de Maringá.

Orientadores: Prof. Dr. Luís Augusto Consularo  
Prof. Dr. Maurício F. Figueiredo

CURITIBA

2003



Ministério da Educação  
Universidade Federal do Paraná  
Mestrado em Informática



## PARECER

Nós, abaixo assinados, membros da Banca Examinadora da defesa de Dissertação de Mestrado em Informática, do aluno Fabrício Martins Lopes, avaliamos o trabalho intitulado, "Um Modelo Perceptivo de Limiarização de Imagens Digitais", cuja defesa foi realizada no dia 22 de agosto de 2003, às dez e trinta horas, no Anfiteatro B do Setor de Ciências Exatas da Universidade Federal do Paraná. Após a avaliação, decidimos pela aprovação do candidato. (Convênio número 279-00/UFPR de Pós-Graduação entre a UFPR e a UEM - ref. UEM número 1331/2000-UEM).

Curitiba, 22 de agosto de 2003.

Prof. Dr. Luís Augusto Consularo  
FCMNTI/UNIMEP - Orientador

Prof. Dr. Mauricio Fernandes Figueiredo  
DIN/UEM - Orientador

Prof. Dra. Júlia Sawaki Tanaka  
IQ/UNESP - Membro Externo

Prof. Dr. Hélio Pedrini  
DINF/UFPR - Membro Interno

## **AGRADECIMENTOS**

A DEUS por iluminar meus caminhos.

Aos meus pais pelo incentivo e amor em todos os momentos.

Ao meu irmão pelas constantes orações e apoio.

A minha namorada Daniela pelo seu carinho e compreensão pela minha constante ausência.

Aos amigos Osvaldo, Nicholas e Lígia pelo companheirismo e tempo que passamos juntos.

Aos colegas do CEFET pelas contribuições neste trabalho, em especial ao amigo Gabriel pelos conselhos e apoio.

Ao meu orientador Prof. Dr. Luís Augusto Consularo pela paciência e generosidade em ensinar, além da sua inestimável amizade.

Ao Prof. Dr. Maurício F. Figueiredo pela sua colaboração neste trabalho.

## SUMÁRIO

AGRADECIMENTOS.....	i
SUMÁRIO .....	ii
LISTA DE FIGURAS .....	iv
LISTA DE TABELAS .....	viii
LISTA DE ALGORITMOS.....	x
RESUMO .....	xi
ABSTRACT .....	xii
1 INTRODUÇÃO.....	1
2 REVISÃO BIBLIOGRÁFICA.....	4
2.1 Atributos das Imagens .....	4
2.1.1. Entropia .....	5
2.1.2. Parâmetros Estatísticos do Contraste em Imagens Digitais .....	7
2.1.3. Parâmetros Estatísticos de Imagens Digitais.....	9
2.2 Limiarização .....	10
2.2.1. Método Otsu.....	15
2.2.2. Análise da Concavidade do Histograma .....	17
2.2.3. Métodos Entrópicos Unidimensionais .....	17
2.2.4. Métodos Entrópicos Bidimensionais .....	25
2.2.5. Método Transformação do Histograma .....	27
2.2.6. Métodos de Segunda Ordem dos Níveis de Intensidade.....	28
2.2.7. Métodos com Abordagens Evolucionárias .....	30
2.2.8. Métodos com Lógica Nebulosa.....	31
2.2.9. Método Aprendizado por Reforço .....	36
2.3 Algoritmos Básicos .....	38
2.3.1. Histograma de Níveis de Intensidade.....	38
2.3.2. Equalização do Histograma.....	39
2.3.3. Exemplo de Limiarização .....	42



2.3.4. Um exemplo de limiarização automática: Máxima Soma das Entropias .....	43
2.3.5. Algoritmo de Agrupamento K-médias.....	44
2.4 Funções de Base Radial (FBR) e Redes de FBRs.....	46
2.5 Psicofísica.....	53
3 METODOLOGIA .....	56
3.1.A Definição de Características .....	56
3.2.A Elaboração do Experimento Psicofísico.....	56
3.3.A Elaboração do Modelo.....	60
3.3.1. A Interpolação Usando Thin-Plate Splines.....	61
3.3.2. Generalização com RFBR.....	69
3.4.Validação da Aplicação do Modelo .....	74
3.4.1. Validação Cruzada Generalizada .....	75
3.4.2. Validação Deixa-Um-Fora .....	77
3.4.3. Validação Bootstrap .....	78
3.5.Comparação entre Modelos de Limiarização.....	79
4 RESULTADOS .....	82
4.1.Resultados Iniciais.....	83
4.2.Resultados intermediários .....	86
4.3.Resultados Finais.....	88
4.4.Comparação entre Limiares Automáticos e Humanos.....	94
4.4.1. Comparação entre Limiares: Modelo Kapur e Humanos.....	95
4.4.2. Comparação entre Limiares: Método Brink e Humanos .....	97
4.4.3. Comparação entre Limiares: Método Otsu e Humanos .....	98
4.5.Comparação entre Limiares RFBR e Humanos .....	100
5 CONCLUSÃO.....	104
6 REFERÊNCIAS .....	106
7 APÊNDICE .....	111
7.1.Imagens Variadas .....	111
7.2.Imagens de Faces.....	114

## LISTA DE FIGURAS

Figura 1 - Exemplo de imagens limiarizadas por humanos e modelo automático.....	2
Figura 2 - Imagens de tamanho (256x256) e suas respectivas entropias.....	6
Figura 3 - Imagem e seu Histograma de Contraste.....	8
Figura 4 - Imagens de tamanho (256x256), suas respectivas variâncias e desvios padrões .....	10
Figura 5 - Histograma de uma imagem em níveis de cinza com limiar global .....	12
Figura 6 - Exemplo de Limiarização multiníveis (60, 85, 140, 177, 230).....	13
Figura 7 - Vizinhança 3x3 de um pixel na coordenada (i,j) da imagem.....	22
Figura 8 - Superfície gerada pelo histograma bidimensional .....	26
Figura 9 - Funções de pertinência nebulosas: S e Z .....	33
Figura 10 - Funções de pertinência por retas .....	34
Figura 11 - Modelo básico de aprendizado por reforço.....	37
Figura 12 - Imagem e seu Histograma.....	38
Figura 13 - Imagem original, FDA e seu Histograma .....	40
Figura 14 - Imagem equalizada, FDA e seu Histograma.....	41
Figura 15 - Limiarização de Imagens .....	42
Figura 16 - Exemplos de funções radiais.....	47
Figura 17 - Rede de funções de base radial tradicional [HAYKIN, 2001].....	51
Figura 18 - Lei de Weber e Fechner .....	54
Figura 19 - Tela inicial do experimento. Nesta tela são mostradas as instruções para que o sujeito clique inicialmente na imagem (da esquerda) antes de escolher algum limiar.....	58
Figura 20 - Captura da janela com os componentes do aplicativo que implementa o experimento. A imagem original (1), a imagem limiarizada (2) pelo limiar escolhido pelo sujeito usando a barra deslizante (3). As caixas de entrada para identificação	

(4), as configurações da borracha para remover pixels que o sujeito avalie não pertencer ao objeto e dois botões: um para gravar o resultado e outro para reiniciar o experimento para aquela imagem (5) .....	59
Figura 21 - Exemplo de Modelo de Limiarização usando a interpolação por base radial com TPS, Cada centro (imagem) está representado por um círculo numerado .....	63
Figura 22 - Aplicação do Modelo de Limiarização TPS. As respostas generalizadas pelo modelo estão representadas por triângulos numerados .....	66
Figura 23 - Geração do modelo usando 3 centros. As cruzes representam as imagens usadas no treinamento e os círculos representam os centros selecionados pelo algoritmo k-médias.....	71
Figura 24 - Geração do modelo usando 10 centros. As cruzes representam as imagens usadas no treinamento e os círculos representam os centros selecionados pelo algoritmo k-médias.....	72
Figura 25 - Aplicação do modelo usando 3 centros. Os triângulos representam as imagens de teste e os asteriscos representam as respostas do modelo para cada uma das imagens .....	73
Figura 26 - Aplicação do modelo usando 10 centros. Os triângulos representam as imagens de teste e os asteriscos representam as respostas do modelo para cada uma das imagens .....	74
Figura 27 - Exemplo de divisão das amostras usadas pela validação cruzada generalizada.....	76
Figura 28 - Exemplo de divisão das amostras usadas pela validação deixa um fora .....	77
Figura 29 - Exemplo de divisão das amostras usadas pela validação bootstrap.....	79
Figura 30 - Software desenvolvido para comparação entre limiares.....	81
Figura 31 - Experimento com melhor desempenho na fase inicial.....	84

Figura 32 - Experimento com pior desempenho na fase inicial .....	85
Figura 33 - Modelo de limiarização usando os atributos: entropia, variância e os limiares humanos com 13 funções de base radial (FBR) .....	90
Figura 34 - Exemplo de imagem usada no conjunto de faces .....	92
Figura 35 - Diagrama de dispersão entre limiares humanos e automáticos propostos por Kapur, aplicados ao conjunto de imagens variadas .....	95
Figura 36 - Diagrama de dispersão entre limiares humanos e automáticos propostos por Kapur, aplicados ao conjunto de imagens de faces .....	96
Figura 37 - Diagrama de dispersão entre limiares humanos e automáticos propostos por Brink, aplicados ao conjunto de imagens variadas .....	97
Figura 38 - Diagrama de dispersão entre limiares humanos e automáticos propostos por Brink, aplicados ao conjunto de imagens de faces .....	98
Figura 39 - Diagrama de dispersão entre limiares humanos e automáticos propostos por Otsu, aplicados ao conjunto de imagens variadas .....	99
Figura 40 - Diagrama de dispersão entre limiares humanos e automáticos propostos por Otsu, aplicados ao conjunto de imagens de faces .....	100
Figura 41 - Diagrama de dispersão entre limiares humanos e melhores limiares automáticos (medianos) do modelo RFBR usando como atributo apenas a entropia, aplicados ao conjunto de imagens variadas .....	101
Figura 42 - Diagrama de dispersão entre limiares humanos e piores limiares automáticos (medianos) do modelo RFBR usando	

como atributo apenas a entropia, aplicados ao conjunto de imagens variadas.....	102
Figura 43 - Diagrama de dispersão entre limiares humanos e melhores limiares automáticos (medianos) do modelo RFBR usando como atributo apenas a variância, aplicados ao conjunto de imagens variadas.....	103
Figura 44 - Diagrama de dispersão entre limiares humanos e piores limiares automáticos (medianos) do modelo RFBR usando como atributo apenas a variância, aplicados ao conjunto de imagens variadas.....	103

## LISTA DE TABELAS

Tabela 1 - Exemplo de imagens usadas durante a geração do modelo TPS e suas respectivas entropias e limiares humanos .....	65
Tabela 2 - Exemplo de imagens teste e suas respectivas entropias, limiares humanos e limiares retornados pelo modelo .....	68
Tabela 3 - Resultados da validação do modelo TPS feito com alguns testes simples, usando como atributo apenas a entropia das imagens .....	84
Tabela 4 - Resultados da validação unidimensional (entropia), que representa a média entre 10 execuções do modelo para os comportamentos conhecidos.....	85
Tabela 5 - Resultados da validação bidimensional (entropia e variância), que representa a média entre 10 execuções do modelo para os comportamentos conhecidos.....	86
Tabela 6 - Resultados da validação do modelo TPS feito com alguns testes simples, usando como atributos a entropia e desvio padrão das imagens.....	87
Tabela 7 - Resultados da validação bidimensional (entropia e variância) do modelo para os comportamentos conhecidos. Os erros RMS representam a média entre 96 execuções para a VCG e VDF e 40 execuções para o bootstrap .....	89
Tabela 8 - Resultados da validação bidimensional (entropia e variância) do modelo para os comportamentos conhecidos. Os erros RMS representam a média entre 88 execuções para a VCG e VDF e 40 execuções para o bootstrap .....	90
Tabela 9 - Resultados da validação aplicada a vários atributos e suas combinações ao modelo para o conjunto de imagens variadas. Os erros RMS representam a média entre 88 execuções para a VCG e VDF e 40 execuções para o bootstrap. Os atributos	

estão codificados, iniciando pelo número que significa a quantidade de atributos combinados e seguido das seguintes abreviações: e = entropia, dp = desvio padrão, im = intensidade média, mev4 = mediana do histograma de contraste, q1v4 = primeiro quartil do histograma de contraste, q3v4 = terceiro quartil do histograma de contraste e maxv4 = máximo do histograma de contraste ..... 91

Tabela 10 - Resultados da validação aplicada a vários atributos e suas combinações ao modelo para o conjunto de imagens de faces. Os erros RMS representam a média entre 45 execuções para a VCG e VDF e 40 execuções para o bootstrap. Os atributos estão codificados, iniciando pelo número que significa a quantidade de atributos combinados e seguido das seguintes abreviações: e = entropia, dp = desvio padrão, im = intensidade média, mev4 = mediana do histograma de contraste, q1v4 = primeiro quartil do histograma de contraste, q3v4 = terceiro quartil do histograma de contraste e maxv4 = máximo do histograma de contraste ..... 93

Tabela 11 - Conjunto de imagens de faces, seus atributos e limiares ..... 113

Tabela 12 - Conjunto de imagens de faces, seus atributos e limiares ..... 115

## LISTA DE ALGORITMOS

Algoritmo 1 - Histograma de uma imagem .....	39
Algoritmo 2 - Equalização de uma imagem .....	41
Algoritmo 3 - Limiarização global genérico.....	43
Algoritmo 4 - Limiar por máxima entropia .....	44
Algoritmo 5 - Algoritmo de agrupamento K-médias.....	45



## RESUMO

O problema de segmentação de imagens digitais tem gerado uma grande diversidade de abordagens para solucioná-lo. Uma abordagem bastante simples, a limiarização, ainda tem sido intensivamente utilizada, por viabilizar aplicações de visão em tempo-real. No entanto, muitas vezes os critérios utilizados na limiarização dependem de atributos entrópicos ou estatísticos da imagem.

Procurando investigar qual a relação destes atributos com respostas subjetivas de limiarização, este trabalho desenvolve um modelo de limiarização de imagens baseado nestas respostas, generalizando-as com um modelo resultante do treinamento de uma Rede de Funções de Base Radial (RFBR). As respostas subjetivas foram colhidas em um experimento psicofísico no qual os sujeitos são instados a escolher um limiar que separe melhor o(s) objeto(s) do fundo de uma imagem digital.

Este trabalho sugere também uma avaliação de métodos automáticos de limiarização comparando as respostas destes métodos com as respostas atribuídas por sujeitos humanos. Este resultado é confrontado com as comparações entre as respostas dos humanos e as respostas do modelo RFBR, apontando para um possível grau de incerteza com que as respostas são geradas em cada método, e que permite seu uso como uma medida de desempenho de métodos automáticos.

## **ABSTRACT**

The digital image segmentation challenge has been generating a plethora of methods and approaches. A quite simple approach, the thresholding, has still been intensively applied to enable real-time vision applications. However, the threshold criteria often depend on entropic or statistical image features.

This work searches a relationship between these features and the subjective choices given by human subjects. So, an image thresholding model based on these subjective choices was developed by training a Radial Basis Functions Network (RBFN). It was implemented on web a psychophysical test on web whose purpose was to ask the subjects for an gray level threshold to give the best division between objects and background.

This work also suggests a measure to evaluate automatic thresholding methods, whose answers were compared to subjective choices. Furthermore, the RBFN-modeled answers were compared to the same subjective choices. Both comparisons indicate the application of an uncertainty degree to evaluate automatic methods.

## 1 INTRODUÇÃO

A tarefa de distinguir objetos em uma imagem, conhecida comumente como segmentação de imagens, pode ser extremamente sofisticada, exigindo filtros bastante elaborados e chegando até a resultados bastante eficientes. Contudo, exigir uma alta qualidade da segmentação, geralmente, exige também um preço alto, isto é, normalmente estes algoritmos apresentam, ou grande complexidade, ou alta especificidade, ou então pouca robustez.

Para conseguir atender a exigências de tempo real, sistemas de processamento digital de imagens se utilizam, em um estágio primário, da limiarização para destacar o objeto do fundo.

A limiarização consiste na identificação, em uma imagem, de um limiar de intensidade em que o objeto melhor se distinga do fundo. A escolha deste limiar pode assumir um critério subjetivo de um operador humano. Contudo, já foram propostos diversos métodos para que esta escolha se dê automaticamente, desde baseados em critérios entrópicos ou estatísticos da imagem [ABUTALEB, 1989], [AHUJA & ROSENFELD, 1978], [BEGHDADI et. al., 1995], [BRINK, 1996], [JOHANNSEN & BILLE, 1982], [KAPUR et. al., 1985], [OTSU, 1978] e [PUN, 1980] até os que utilizam algum mecanismo de busca mais sofisticado [CHENG & CHEN, 1999], [CHUN & YANG, 1996], [JAWAHAR et. al., 2000], [YIN, 1999] e [YIN, 2002]. Percebe-se nos resultados destes métodos uma boa qualidade de segmentação, tendo como critério de qualidade uma opinião subjetiva e que, algumas vezes, é adotada como parâmetro para avaliação do método [BEGHDADI et. al., 1995], [BRINK, 1996], [BRINK & PENDOCK, 1996], [BRINK, 1992], [YANOWITZ & BRUCKSTEIN, 1989] e [YIN, 2002]. Para ilustrar, a Figura 1 mostra duas imagens limiarizadas por um método automático e por um sujeito humano, mostrando dois tipos de resultado, um muito bom (Figura 1.a e Figura 1.b) e outro muito ruim (Figura 1.c e Figura 1.d).



(a) Imagem limiarizada por humano, limiar = 126.



(b) Imagem limiarizada pelo método de Brink [BRINK, 1996], limiar = 126.



(c) Imagem limiarizada por humano, limiar = 174.



(d) Imagem limiarizada pelo método de Brink [BRINK, 1996], limiar = 144.

Figura 1 - Exemplo de imagens limiarizadas por humanos e modelo automático.

Embora existam medidas quantitativas da qualidade dos limiarizadores automáticos, estas são geralmente baseadas em propriedades intrínsecas das imagens (forma [SAHOO et. al., 1988] e uniformidade [LEVINE & NAZIF, 1985]), ignorando os motivos da limiarização. É comum encontrar trabalhos nos quais se lança mão da avaliação subjetiva do método, como citado a pouco. No entanto, se estes métodos automáticos podem ser subjetivamente tão bem avaliados para limiarização automática, qual é a relação que existe entre os critérios adotados por seus algoritmos e a percepção visual humana?

Este trabalho responde algumas destas questões apresentando um modelo de percepção visual de limiarização utilizando generalização por Rede de Funções de

Base Radial (RFBR). Os resultados foram validados com três métodos que oferecem dois tipos de avaliação, uma pessimista e outra realista.

As respostas dos métodos automáticos a um conjunto de imagens digitais de motivos variados são comparadas às respostas colhidas de sujeitos humanos, em um experimento psicofísico no qual os sujeitos escolhem limiares de imagens do mesmo conjunto.

Além disso, as respostas do experimento foram submetidas a uma generalização com uma RFBR, que resultou em modelos baseados em atributos entrópicos e estatísticos do contraste das imagens. As melhores e piores respostas destes modelos foram comparadas também com as respostas obtidas dos sujeitos humanos.

Estas comparações sugerem o uso deste tipo de modelo na avaliação de métodos automáticos de limiarização, bem como na construção de um seletor de características para limiarização.

Esta dissertação inicia-se com um capítulo de revisão bibliográfica, o capítulo 2, no qual são apresentados os métodos utilizados ao longo deste trabalho. Nesta revisão constam desde a implementação de métodos básicos como a limiarização, equalização, entre outros e faz também um sobrevôo sobre os métodos automáticos de limiarização. A seguir, no capítulo 3, é apresentada a metodologia utilizada neste trabalho, descrevendo desde os métodos de generalização, até o experimento psicofísico realizado. No capítulo 4 são apresentados os resultados das comparações citadas acima, ilustrando-as com diagramas de dispersão e tabelas. Todas as imagens usadas neste trabalho são apresentadas em um apêndice.

## 2 REVISÃO BIBLIOGRÁFICA

Comparar os resultados de um experimento realizado com sujeitos humanos com métodos automáticos de limiarização requer a introdução de alguns conceitos, métodos e algoritmos.

Este capítulo oferece:

- 1) uma breve revisão sobre os atributos de imagens mais utilizados em métodos de limiarização automática relatados na literatura;
- 2) uma pequena coletânea sobre métodos de limiarização automática apresentando desde o método mais simples (manual) até métodos mais sofisticados e que envolvem aprendizado.
- 3) um conjunto de algoritmos básicos que foram aplicados de alguma forma neste trabalho, para explicitar os métodos utilizados neste trabalho;
- 4) uma explicação didática sobre funções de base radial e seu uso em algoritmos de interpolação e regressão;
- 5) uma revisão sobre métodos psicofísicos;

### 2.1 Atributos das Imagens

Nesta seção são resumidamente descritos alguns atributos de imagens utilizados neste trabalho. Estes atributos são comumente aplicados no reconhecimento de padrões como critérios em classificadores, indicando um grau de separação entre classes ou determinando a pertinência a uma classe. A entropia, em particular, é bastante utilizada na compressão de dados como uma medida de quantidade de informação, servindo como um parâmetro de desempenho deste tipo de algoritmo. O contraste e a intensidade média são atributos bastante simples, porém fundamentais em qualquer método de limiarização. A variância, assim como a entropia, é bastante utilizada nos métodos automáticos de limiarização e também como uma medida de uniformidade e de separação de classes.

### 2.1.1. Entropia

A entropia foi originalmente introduzida por Boltzmann como uma medida macroscópica da organização de sistemas termodinâmicos [BOLTZMANN, 1974]. Mais tarde foi aplicada na Teoria da Informação por Claude Shannon [SHANNON, 1948]. A entropia é freqüentemente utilizada para indicar a quantidade de informação contida em uma determinada fonte. Mas também é utilizada para graduar a desordem (incerteza) de um conjunto de dados [BISHOP, 1995]. Existem diversas definições de entropia, mas a mais difundida na Teoria de Informação cresce de acordo com a presença de mais informação em uma fonte de dados. Em outras palavras, quanto mais incertos forem os símbolos ou padrões em um determinado conjunto de dados, maior a entropia de todo o conjunto.

Uma determinada informação pode ser modelada como um processo probabilístico [GONZALEZ & WOODS, 2000]. Assim, a ocorrência de um evento  $i$  com probabilidade  $p_i$  próxima de um ou de zero, contribuirá para uma entropia mais próxima de zero, ou seja, a incerteza dos eventos é baixa. Isso significa que a informação atribuída ao evento  $i$  é inversamente relacionada à probabilidade de sua ocorrência, ou seja, quanto mais incomum for o evento  $i$ , maior será sua contribuição para o aumento da entropia [STEVENS, 1998].

O conceito de entropia,  $H$ , no processamento de imagens foi introduzido por Pun [PUN, 1980]. Neste caso, considerando a imagem como resultado de um processo aleatório, a probabilidade  $p_i$  na equação (1) corresponde à probabilidade de um pixel em uma imagem digital assumir um valor de intensidade  $i$  ( $i=1,...,G$ ). O histograma dos níveis de intensidade da imagem se transforma facilmente na função densidade de probabilidade (fdp) da imagem, apenas dividindo cada número de pixels de intensidade  $i$ ,  $g_i$ , pelo número de pixels na imagem,  $N$ .

$$H = -\sum_{i=1}^G p_i \log p_i, \quad \sum_{i=1}^G p_i = 1, \quad p_i = \frac{g_i}{N} \quad (1)$$

Sendo:

$g_i$  é a quantidade de pixels com intensidade  $i$ .

$G$  é o número de níveis de cinza na imagem (ou no canal).

$p_i$  representa a probabilidade do nível de cinza  $i$  ser encontrado na imagem.

$N$  é a quantidade total de pixels da imagem (altura x largura).

$H$  é a entropia da imagem.

A entropia de uma imagem é uma medida não negativa e seu resultado, quando a base do logaritmo é dois, é dado em bits [DUDA et. al., 1998]. Dessa forma, o menor valor para a entropia é zero, isso ocorre quando todos os pixels são de uma mesma intensidade  $k$ , isto é, a probabilidade  $p_k = 1$  e  $p_i = 0$  para  $i \neq k$ , deixando de existir incerteza. No outro extremo, a máxima entropia ocorre quando uma imagem contém a mesma quantidade de pixels para todas as intensidades (ruído branco), isto é, todas as intensidades têm a mesma probabilidade de ocorrerem. Um exemplo dos dois extremos podem ser vistos respectivamente nas Figura 2.a e Figura 2.g, bem como exemplos intermediários.

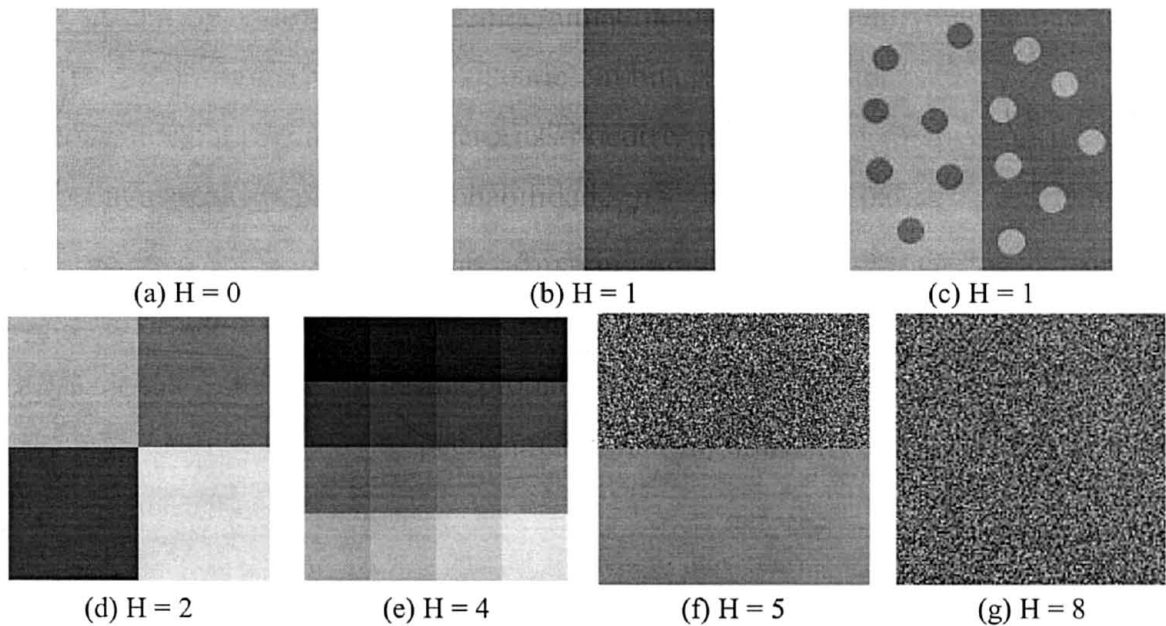


Figura 2 - Imagens de tamanho (256x256) e suas respectivas entropias.



A Figura 2 exemplifica a influência do crescimento do número de intensidades na medida da entropia da imagem. A Figura 2.f e a Figura 2.g possuem todas as 256 intensidades possíveis, mas com probabilidades diferentes, ilustrando que a entropia da imagem não está relacionada apenas com a quantidade de intensidades de cinza, mas também com a probabilidade de suas ocorrências.

Outra observação é que a entropia não está relacionada com a disposição espacial da informação. Exemplos são mostrados na Figura 2.b e na Figura 2.c, que possuem a mesma quantidade de pixels com as mesmas intensidades, porém distribuídos espacialmente de forma diferente e com a mesma entropia. Portanto, a equalização do histograma de uma imagem não altera a quantidade de informação presente nela. A equalização apenas aumenta o contraste da imagem [CASTLEMAN, 1996].

A decisão de uma limiarização pode ser tomada por um critério que não depende da distribuição espacial? Essa pergunta pode ser respondida associando ao critério de limiarização, algum atributo que admita a influência espacial e o contraste é talvez o mais simples deles.

### **2.1.2. Parâmetros Estatísticos do Contraste em Imagens Digitais**

O contraste é uma medida que se refere a diferenças locais de intensidades na imagem. Essas diferenças são observadas na vizinhança de cada pixel. Se a média de intensidades da vizinhança for próxima da intensidade do pixel, conclui-se que o local possui baixo contraste, caso contrário, alto contraste.

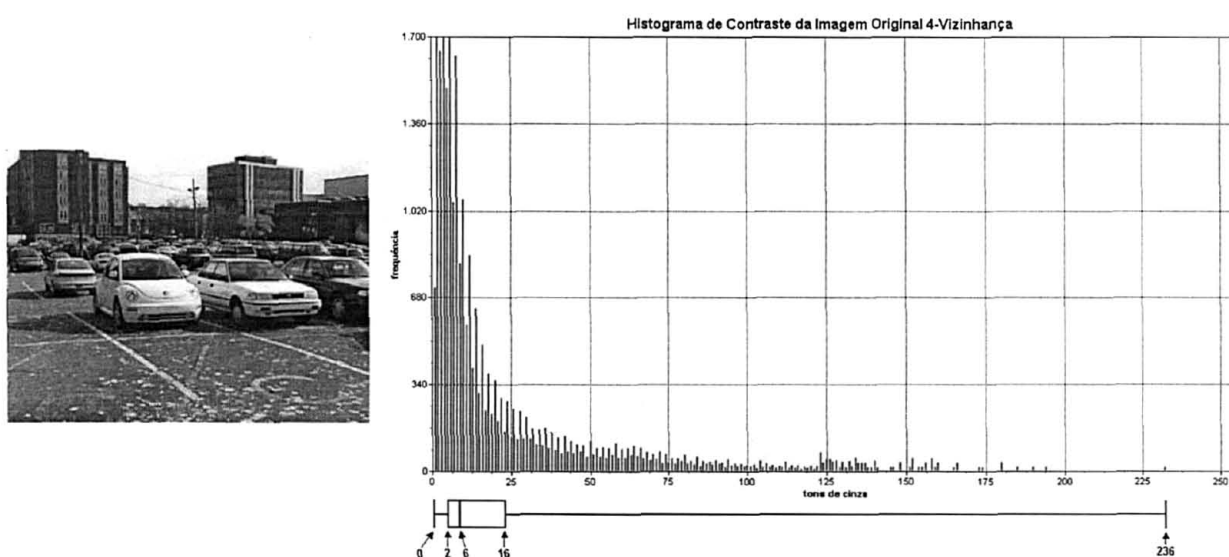
As regiões de baixo contraste podem ser associadas aos objetos ou ao fundo da imagem, por apresentarem regiões homogêneas. Por outro lado, o alto contraste é normalmente detectado nas regiões próximas às bordas, por serem regiões onde existe uma mudança brusca de intensidades.

Neste trabalho foi utilizado o histograma de contraste, calculado sobre uma matriz cujos elementos são formados pela média da soma das diferenças entre a intensidade do pixel e as intensidades de seus pixels vizinhos (em uma 4-vizinhança).

O histograma representa o agrupamento de pixels que possuem a mesma diferença média de intensidades em suas vizinhanças. Um exemplo pode ser observado na Figura 3.

O histograma de contraste caracteriza uma imagem. Contudo, um histograma pode conter muitos valores (o número de diferenças entre a intensidade máxima e mínima) que ainda contenham redundâncias. É conveniente caracterizar o histograma com parâmetros da distribuição observada em uma imagem. Para representar a distribuição, lança-se mão dos cinco parâmetros estatísticos [MOORE, 2000]: mínimo, primeiro quartil, mediana, terceiro quartil e máximo.

Estes parâmetros são ilustrados na Figura 3.b como um retângulo e duas hastes. Em cada extremidade das hastes estão os valores mínimo e máximo (no exemplo, respectivamente os valores 0 e 236). No limite inferior do retângulo está o primeiro quartil (no exemplo, valor 2) e no limite superior o terceiro quartil (no exemplo, valor 16). A mediana (no exemplo, valor 6) aparece como uma linha dividindo o retângulo em dois.



(a) Imagem em tons de cinza.<sup>1</sup>

(b) Histograma de Contraste da Imagem (a)

Figura 3 - Imagem e seu Histograma de Contraste.

<sup>1</sup> VASC. The Vision and Autonomous Systems Center of The Robotics Institute at Carnegie Mellon University (Pittsburg, PA). <http://vasc.ri.cmu.edu/idb/images/car/>

### 2.1.3. Parâmetros Estatísticos de Imagens Digitais

Uma outra maneira simples de caracterizar uma imagem é utilizando seus parâmetros estatísticos. Neste caso, o próprio histograma de intensidades pode ser utilizado para o cálculo da intensidade média (ou primeiro momento) e da variância (ou segundo momento). A média, juntamente com a variância, caracteriza a iluminação durante a aquisição da imagem.

A variância de uma imagem indica a variação das intensidades de cinza em relação à sua média, caracterizando o contraste da imagem. A variância, assim como a entropia, não se altera com a disposição espacial da informação em uma imagem, como ilustra a Figura 4.b e Figura 4.c.

O cálculo da intensidade média e da variância de uma imagem digital é simples, como pode ser visto nas equações (2).a e (2).b, respectivamente.

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1}^L \sum_{j=1}^C g_{ij}}{L \times C} & \sigma^2 &= \frac{\sum_{i=1}^L \sum_{j=1}^C (g_{ij} - \bar{x})^2}{L \times C} \end{aligned} \quad (2)$$

Assumindo:

$L$  representa o número de linhas da imagem.

$C$  representa o número de colunas da imagem.

$g_{ij}$  representa a intensidade presente na coordenada  $(i, j)$  da imagem.

$\bar{x}$  como a intensidade média da imagem.

$\sigma^2$  como a variância da imagem.

A Figura 4 ilustra imagens digitais e suas respectivas variâncias. A Figura 4.a apresenta variância nula, pois todos os pixels têm a mesma intensidade. A Figura 4.b e a Figura 4.c mostram uma variância máxima, pois seus pixels variam, na mesma quantidade, de um extremo a outro da intensidade, não importando a distribuição espacial destes pixels na imagem. A Figura 4.d tem variância média entre seus elementos.

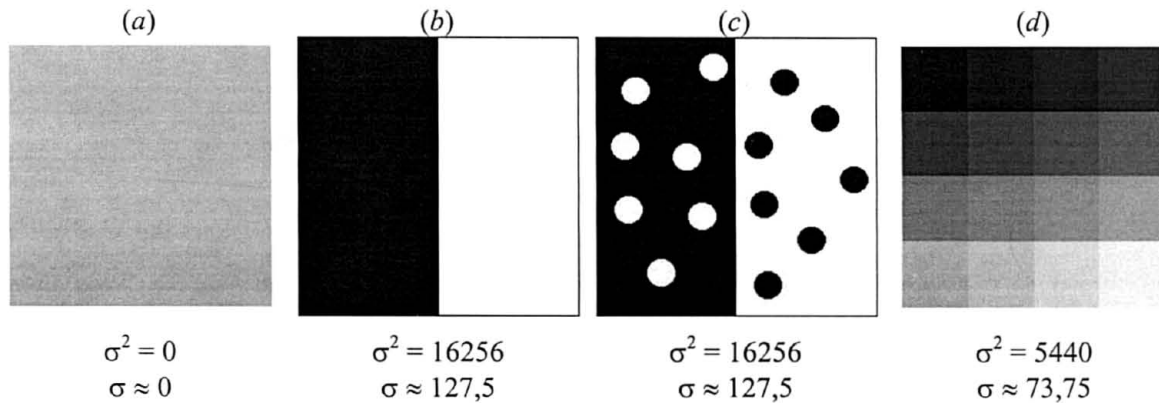


Figura 4 - Imagens de tamanho (256x256), suas respectivas variâncias e desvios padrões.

## 2.2 Limiarização

A limiarização de imagens é uma ferramenta de segmentação muito simples e computacionalmente rápida. É indicada para aplicações que exijam a identificação e extração do objeto da imagem, cujo tempo computacional despendido seja um fator relevante.

As técnicas de segmentação estão divididas em dois conceitos distintos: similaridade (agrupamento de regiões homogêneas) ou descontinuidades (detecção de bordas). A limiarização é uma forma de agrupamento de regiões homogêneas [GONZALEZ & WOODS, 2000].

Alguns pesquisadores procuram classificar as técnicas de limiarização normalmente observando quais informações são usadas na escolha do limiar e o número de limiares utilizados. Weska [WESKA, 1978] classifica os métodos de limiarização em três grupos: global, local e dinâmico. O critério usado nesta definição é a dependência do operador de limiar. Se for determinado usando apenas as intensidades é classificado como global. Se usar as intensidades e alguma informação espacial da imagem é classificado como local. Se usar ambas as informações mais as coordenadas espaciais é classificado como dinâmico.

Sahoo [SAHOO et. al.1988], por sua vez, divide as técnicas em dois grandes grupos: globais e locais. Os limiares globais são aqueles que dividem toda a imagem usando apenas um limiar. Os locais dividem a imagem em subimagens e para cada

subimagem é atribuído um limiar. Esta técnica também é conhecida como limiar multiníveis. Esta classificação usa como critério a quantidade de limiares selecionados para a imagem.

Ainda em seu trabalho, Sahoo classifica os limiares globais e locais como: ponto-dependente ou região-dependente. Se o limiar é determinado apenas pelas intensidades de cada pixel, então o método é chamado ponto-dependente. Se o limiar é determinado por alguma propriedade local como, por exemplo, a distribuição de intensidades da vizinhança de cada pixel, então é região-dependente.

Zhang [ZHANG, 1996] classifica os algoritmos de segmentação de imagens também em duas classes distintas: métodos analíticos e métodos empíricos. Os métodos analíticos são baseados na análise de princípios, requisitos, complexidade e propriedades do algoritmo. Por outro lado, os empíricos são caracterizados apenas pela qualidade de seus resultados.

Neste trabalho é adotada a classificação proposta por Sahoo, a qual apresenta mais critérios de classificação sendo conseqüentemente mais precisa. Sem desprezar a simplicidade, também é a mais comumente usada pelos pesquisadores desta área.

Definidos os limiares globais e multiníveis, a aplicação do limiar global em uma imagem bidimensional  $I(x,y)$  com níveis de intensidade  $N_c = \{0,1,2,\dots,255\}$  consiste em determinar um único limiar  $T \in N_c$  que separe os pixels em duas classes distintas: objeto e fundo. Esta técnica, normalmente é aplicada ao histograma da imagem  $h(N_c)$ , o qual pode ser visto como uma descrição da distribuição das intensidades pela imagem.

Na Figura 5 se vê um histograma com as intensidades variando entre 0 (preto) e 255 (branco). Vê-se também dois picos (histograma bimodal): o primeiro representa o objeto (escuro) e o segundo o fundo da imagem (claro). Eles estão separados por um vale, que neste exemplo seria o local mais apropriado para estabelecer o limiar.

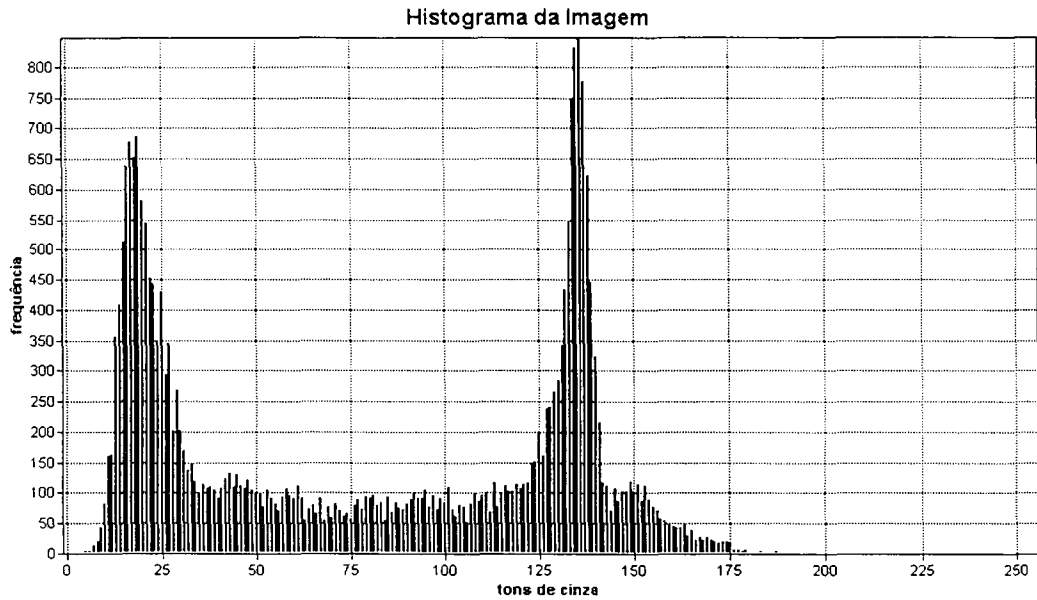


Figura 5 - Histograma de uma imagem em níveis de cinza com limiar global.

Após determinar um limiar  $T$ , a imagem inteira fica dividida em duas regiões: o objeto, que recebe a cor preta (0) e o fundo da imagem, que recebe a cor branca (255), ou vice-versa. Desta forma, uma imagem limiarizada  $LI(x, y)$  pode ser definida pela equação (3), a qual é representada na forma de um algoritmo na seção 2.3.3.

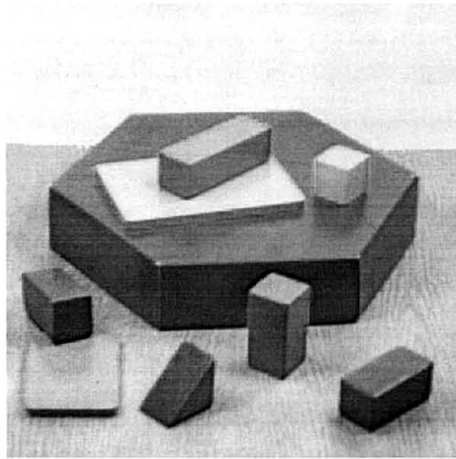
$$LI(x, y) = \begin{cases} 255, & \text{se } I(x, y) > T \\ 0, & \text{se } I(x, y) \leq T \end{cases} \quad (3)$$

O limiar multiníveis também pode ser aplicado ao histograma da imagem. Seu desenvolvimento tem início com a divisão do histograma em sub-regiões, também chamadas de subimagens, as quais representam regiões homogêneas de intensidade. Para cada uma das sub-regiões é determinado um limiar. Assim, o limiar multiníveis pode ser entendido como um conjunto de limiares  $\bar{t} = \{t_1, t_2, t_3, \dots, t_n\}$  com  $2 \leq n \leq 255$ ,  $t_1 < t_2 < t_3 < \dots < t_n$  e o conjunto  $\bar{t} \in C_{Nc}$ .

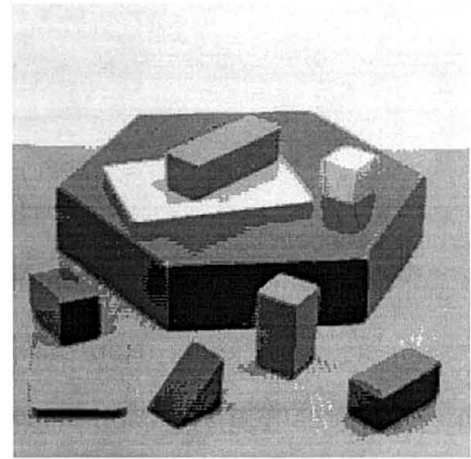
Utilizando a imagem bidimensional em níveis de cinza  $I(x, y)$  como entrada, a imagem multi-limiarizada  $LI_M(x, y)$  é definida na equação (4).

$$LI_M(x, y) = \begin{cases} mc_0, & \text{se } I(x, y) \leq t_1 \\ mc_1, & \text{se } t_1 < I(x, y) \leq t_2 \\ mc_2, & \text{se } t_2 < I(x, y) \leq t_3 \\ \vdots & \vdots \\ mc_n, & \text{se } I(x, y) > t_n \end{cases} \quad (4)$$

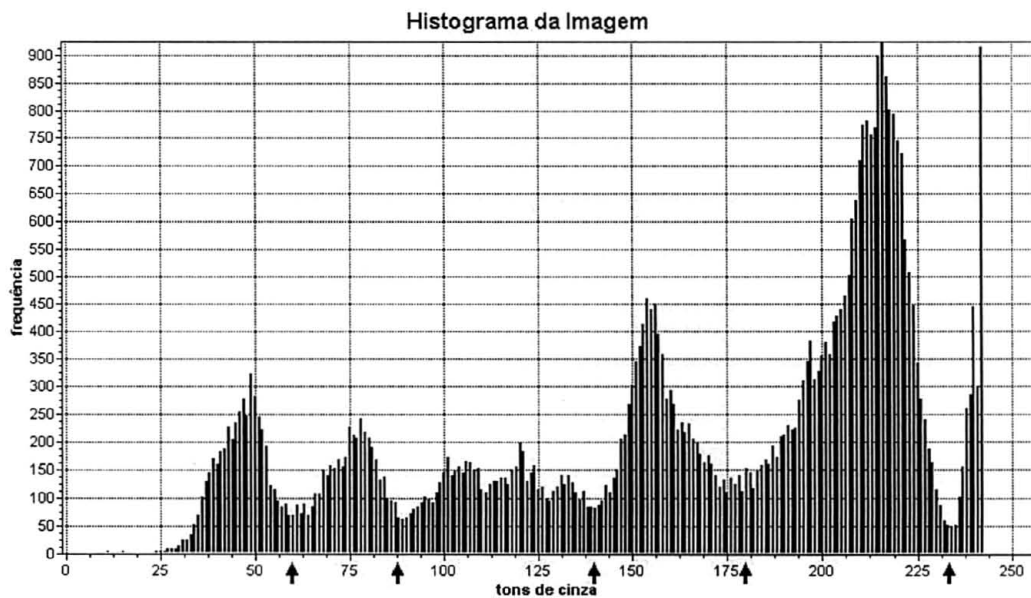
Os valores das intensidades de cinza atribuídos a  $mc_i$  normalmente são fixados como a média das intensidades entre cada intervalo de limiares  $(t_i, t_{i+1})$ , assumindo  $t_0 = 0$ . Para estabelecer os limiares multiníveis duas perguntas precisam ser respondidas: <sup>(1)</sup> “Qual número de limiares deverá ser utilizado?” e <sup>(2)</sup> “Quais os valores adotados em cada limiar?”.



(a) Imagem em tons de cinza



(b) Imagem limiarizada multiníveis



(c) Histograma da imagem (a)

Figura 6 - Exemplo de Limiarização multiníveis (60, 85, 140, 177, 230).

As respostas para estas perguntas variam de acordo com as características apresentadas em cada imagem. Na Figura 6.c, vê-se um histograma com vários picos (multimodal), os quais normalmente representam objetos distintos da imagem separados por pequenos vales. Analisando este exemplo, os valores mais apropriados para a quantidade de limiares e o valor assumido por cada limiar seriam respectivamente  $n = 5$  e  $\bar{t} = \{60, 85, 140, 177, 230\}$ .

Segundo González [GONZALEZ & WOODS, 2000], a limiarização multiníveis é geralmente mais trabalhosa que o limiar global, pela dificuldade em encontrar múltiplos limiares que determinem efetivamente as regiões de interesse, especialmente quando houver vários grupos de objetos na imagem.

Por outro lado, a limiarização global alcança bons resultados apenas quando a iluminação da imagem é relativamente uniforme e as regiões de interesse, representadas pelos objetos, possuam uma diferença significativa de intensidade do seu fundo (contraste).

De qualquer forma, o uso da limiarização multiníveis aumenta consideravelmente o tempo computacional despendido na determinação dos limiares. É mais trabalhoso e pode gerar resultados tão simples ou similares quanto os resultados obtidos com limiares globais.

Assim, o limiar global se tornou uma ferramenta automática amplamente usada nas mais diversas áreas e atualmente reúne muitos pesquisadores interessados em, cada vez mais, melhorar seu desempenho. Muitos métodos de limiarização têm sido propostos, além de variações destes métodos. Muitos deles podem ser encontrados nos trabalhos de Fu e Mui [FU & MUI, 1981], de Sahoo [SAHOO et al., 1988], de Lee [LEE et al., 1990] e de Pal [PAL & PAL, 1993].

Um dos primeiros métodos de limiarização automáticos é o *P-tile*, ele foi proposto por Doyle em 1962 [DOYLE, 1962]. Neste método os objetos são assumidos como escuros com um fundo claro ou vice-versa. A porcentagem que o objeto ocupa da imagem é um dado conhecido. Dessa forma, o valor do limiar é determinado respeitando o número de pixels ocupados pelo objeto. Por exemplo, se o objeto



representa 25% da imagem, o limiar é determinado pelo valor que divida aproximadamente 25% dos pixels como sendo objeto e 75% como sendo fundo.

Como se pode observar, sua limitação é que só é aplicado em imagens cuja porcentagem da área ocupada pelo objeto é conhecida, sua implementação é sequencial e seleciona apenas um limiar para toda a imagem, ou seja, um limiar global.

Em seguida em 1966, Prewit e Mendelsohn [PREWIT & MENDELSON, 1966] propuseram o método *Mode*. Este método é usado apenas em imagens nas quais o objeto e fundo sejam distintos, isto é, aquelas imagens cujo histograma é bimodal. Observando esta restrição, a escolha do limiar é feita entre os picos do histograma, onde é procurada a intensidade mais profunda do vale. Este método é um dos mais simples, sua implementação é sequencial e o limiar é global.

Nas próximas subseções serão apresentadas resumidamente algumas técnicas automáticas de limiarização de imagens e suas variações, bem como algumas vantagens e desvantagens. É importante destacar que dois critérios são comumente usados para avaliar o desempenho dos métodos de limiarização: uniformidade e forma.

Proposta por Levine e Nazif [LEVINE & NAZIF, 1985], a uniformidade é inversamente proporcional à variância das intensidades em cada região, isto é, objeto e fundo da imagem. Isto significa que, se a região do objeto possuir variância alta, será classificada como baixa uniformidade. O critério de forma foi proposta por Sahoo [SAHOO et al., 1988] e está relacionado com a identificação da forma do objeto na imagem em relação ao seu fundo.

### 2.2.1. Método Otsu

Este método foi proposto por Otsu [OTSU, 1978]. Sua principal característica é a maximização da variância entre as classes da imagem, isto é, a maximização da separação entre o objeto e o fundo da imagem.

A operação de limiar é vista como o particionamento dos pixels de uma imagem em duas classes:  $C_O$  (objeto) e  $C_F$  (fundo).  $C_O = \{0, 1, \dots, T\}$  e

$C_F = \{T + 1, T + 2, \dots, G - 1\}$ . Sendo:  $T$  o limiar escolhido e  $G$  o número de níveis de intensidade da imagem.

O limiar ótimo é determinado pelo valor que maximize um dos seguintes fatores.

$$\lambda = \frac{\sigma_B^2}{\sigma_W^2} \quad \eta = \frac{\sigma_B^2}{\sigma_T^2} \quad \kappa = \frac{\sigma_T^2}{\sigma_W^2} \quad (5)$$

Segundo Sahoo [SAHOO et al., 1988], das três equações descritas em (5), por ser usada, por exemplo,  $\eta$ . Assim, o limiar ótimo pode ser determinado pela equação (6).

$$T = \underset{T \in G}{\text{Max}}(\eta), \quad (6)$$

Sendo:

$$\begin{aligned} \sigma_T^2 &= \sum_{i=0}^G (i - \mu_T)^2 p_i & \mu_T &= \sum_{i=0}^G i p_i & \sigma_B^2 &= \omega_0 \omega_1 (\mu_1 - \mu_0)^2 \\ \omega_0 &= \sum_{i=0}^T p_i & \omega_1 &= 1 - \omega_0 & \mu_1 &= \frac{\mu_T - \mu_0}{1 - \omega_0} \\ \mu_0 &= \frac{\mu_S}{\omega_0} & \mu_S &= \sum_{i=0}^T i p_i & \sum_{i=1}^G p_i &= 1, \quad p_i = \frac{g_i}{N} \end{aligned}$$

$g_i$  representa a quantidade de pixels com intensidade  $i$ .

$G$  representa o número de intensidades de cinza da imagem.

$p_i$  representa a probabilidade do nível de intensidade  $i$  ser encontrado na imagem.

$N$  representa a quantidade total de pixels da imagem.

Uma das vantagens deste método é que ele não se restringe ao tipo de histograma da imagem. Isto é, pode ser aplicado em histogramas unimodais, bimodais ou multimodais, mas possui melhor desempenho em imagens com maior variância de intensidade [OTSU, 1978].

Considerando as medidas de desempenho uniformidade e forma, este método apresenta bons resultados para imagens com histogramas bimodais e unimodais. É classificado por Sahoo e Lee, como um método orientado à uniformidade do objeto [SAHOO et. al., 1988] [LEE et. al., 1990].

Sua principal desvantagem, segundo Lee [LEE et. al., 1990], está no desempenho sensível à presença de ruídos na imagem. Por outro lado, ele considera este método uma das melhores escolhas para aplicações gerais em tempo-real em visão computacional. Sua implementação é seqüencial e determina um limiar global.

### **2.2.2. Análise da Concavidade do Histograma**

Este método foi proposto por Rosenfeld e De La Torre em 1983 [ROSENFELD & DE LA TORRE, 1983]. A principal aplicação deste método é em imagens cujos histogramas não apresentam vales bem definidos, chegando a atribuir bons limiares nos picos do histograma, desde que os vales e os picos formem uma concavidade, e assim, o limiar pode ser determinado pela análise dessa concavidade do histograma.

Uma vantagem é que esse método pode ser aplicado a histogramas unimodais, mas desde que apresentem um mínimo de concavidade entre os picos. Por outro lado, segundo [SAHOO et. al., 1988], se for considerada a uniformidade e a forma da imagem limiarizada, em geral este método não produz bons resultados em comparação com outros métodos de limiarização. Sua implementação é seqüencial e determina um limiar global.

### **2.2.3. Métodos Entrópicos Unidimensionais**

Este método é um dos mais pesquisados e que recebeu vários incrementos ao longo do tempo. Por apresentar várias abordagens, foram divididos em duas seções: unidimensionais e bidimensionais. Esta classificação considera a dimensão das informações processadas na determinação do limiar.

Em 1980 com o trabalho de Pun [PUN, 1980], tem início um novo tipo de método de limiarização: os métodos entrópicos. Como o próprio nome diz, é baseado na entropia da teoria da informação, mais especificamente na análise da informação contida no histograma da imagem.

Este método tem como base a maximização da informação (entropia, seção 2.1.1) da imagem, dividindo o seu histograma em duas distribuições de probabilidade, uma entre os valores de 0 a  $T$  e outra para os valores de  $T + 1$  até  $G$ , representando as duas classes presentes na imagem (objeto e fundo). Estas duas distribuições inicialmente propostas por Pun [PUN, 1980] são dadas por:

$$H_b(T) = -\sum_{i=0}^T p_i \log_e p_i \quad H_w(T) = -\sum_{i=T+1}^G p_i \log_e p_i \quad \sum_{i=1}^G p_i = 1, \quad p_i = \frac{g_i}{N} \quad (7)$$

Sendo:

$g_i$  representa a quantidade de pixels com intensidade  $i$ .

$G$  representa o número de intensidades de cinza da imagem.

$p_i$  representa a probabilidade do nível de cinza  $i$  ser encontrado na imagem.

$N$  representa a quantidade total de pixels da imagem.

$H_b$  e  $H_w$  armazenam informações associadas, respectivamente, aos pixels pretos e brancos, após a limiarização. A escala de níveis de intensidade é percorrida com o objetivo de encontrar um valor de  $T$  (limiar) que maximize a função da equação (8), e assim, encontrar a intensidade que melhor separe o objeto e o fundo da imagem.

$$\psi(T) = \text{MAX} \{ H_b(T) + H_w(T) \} \quad (8)$$

O limiar resultante representa a melhor divisão entre o objeto e fundo da imagem de acordo com a máxima soma das entropias.

Em 1982, Johannsen e Bille [JOHANNSEN & BILLE, 1982] propuseram uma nova forma de usar a entropia do histograma para determinar o limiar. Como no método anterior, o histograma continua sendo dividido em duas partes, mas agora considerando o limiar que minimize a interdependência entre as regiões (objeto e fundo). Assim, as entropias passaram a ser definidas pela equação (9).

$$\begin{aligned}
H_b(T) &= \log_e \left( \sum_{i=0}^T p_i \right) - \frac{1}{\sum_{i=0}^T p_i} \left[ p_T \log_e p_T + \left( \sum_{i=0}^{T-1} p_i \right) \log_e \left( \sum_{i=0}^{T-1} p_i \right) \right], \\
H_w(T) &= \log_e \left( \sum_{i=T}^G p_i \right) - \frac{1}{\sum_{i=T}^G p_i} \left[ p_T \log_e p_T + \left( \sum_{i=T+1}^G p_i \right) \log_e \left( \sum_{i=T+1}^G p_i \right) \right] \\
p_T &= \sum_{i=0}^T p_i
\end{aligned} \tag{9}$$

Dadas as duas entropias  $H_b(T)$  e  $H_w(T)$ , o processo de busca é o mesmo anteriormente apresentado, o qual executa um processo de busca exaustiva pelas intensidades com o objetivo de encontrar aquela intensidade que, agora, minimize a soma das entropias, como mostra a equação (10).

$$T = \text{Min}\{H_b(T) + H_w(T)\} \tag{10}$$

Pouco tempo depois, em 1985, Kapur [KAPUR et al., 1985] propôs que o cálculo das entropias do objeto e do fundo, fosse alterado, passando a ser obtidos como em (11).

$$H_b(T) = - \sum_{i=1}^T \frac{p_i}{P_0} \log \frac{p_i}{P_0}, \quad H_w(T) = - \sum_{i=T+1}^G \frac{p_i}{P_1} \log \frac{p_i}{P_1} \tag{11}$$

Sendo:

$$P_0 = \sum_{i=1}^T p_i, \quad P_1 = \sum_{i=T+1}^G p_i, \quad P_0 + P_1 = 1.$$

$p_i$  representa a probabilidade do nível de cinza  $i$  ser encontrado na imagem.

$G$  representa o número de intensidades de cinza da imagem.

$H$  representa a entropia da imagem.

Dadas as duas entropias  $H_b(T)$  e  $H_w(T)$ , é executado um processo de busca exaustiva pelas intensidades, com o objetivo de encontrar a intensidade que maximize a equação (8).

É importante observar que as abordagens propostas por Pun, Kapur e Johannsen e Bille não consideram a correlação espacial entre os pixels de uma imagem, ou seja, assume-se que as intensidades dos pixels sejam independentes entre si. Com esta independência a qualidade pode ficar comprometida.

Conforme apresentado na seção 2.1.1, a entropia não está relacionada com a disposição espacial dos pixels na imagem, o que significa que os pixels poderiam ser reorganizados de qualquer outra forma, e o resultado permaneceria o mesmo.

Buscando melhorar o desempenho, Beghdadi [BEGHDADI et. al., 1995] propôs em seu trabalho que a imagem fosse dividida em blocos quadrados (janelas) de tamanho  $s$  (maior que um pixel), com a intenção de identificar possíveis padrões na imagem, sendo esta a fonte de informação (entropia) considerada. Segundo Beghdadi, o tamanho das janelas pode ser fixo ou variável.

Para implementar esta abordagem gera-se para cada bloco, o seu respectivo bloco binarizado por um limiar  $T$ . Este procedimento cria um conjunto  $B$  de blocos binarizados com tamanho  $N = 2^{sx_s}$ . Para encontrar a distribuição de probabilidade é criado um subconjunto  $B_k$  contendo todos os blocos com  $k$  pixels brancos, assim a probabilidade de encontrar um bloco  $P_k$  com  $k$  pixels brancos é:

$$P_k = \frac{N(B_k)}{N} \quad (12)$$

Tendo gerado todas as combinações para todos os blocos da imagem, o procedimento de busca pelo limiar consiste em encontrar o valor máximo para a entropia  $H(T)$  como mostra (13).

$$H(T) = -\sum_{k=1}^N P_k(t) \log P_k(t) \quad T = \text{Arg}\{\text{Max}[H(T)]\} \quad (13)$$

Isto significa que a intensidade  $T$  que gerar a maior quantidade de blocos binarizados iguais (quantidade de pixels brancos ou pretos), sem considerar a posição espacial dos pixels dentro do bloco, terá a maior probabilidade de ocorrer e, conseqüentemente, será o limiar escolhido.

Uma vantagem deste método é considerar a informação espacial dos pixels dentro de cada bloco, mas cada um deles é assumido como sendo independente.

O tamanho da janela é uma das desvantagens do método, uma vez que pode ser muito pequeno e assim insuficiente para descrever o padrão contido na imagem ou pode ser muito grande associando padrões locais desconexos.

Pouco tempo depois, Brink e Pendock [BRINK & PENDOCK, 1996], desenvolveram uma técnica de limiarização baseada na entropia, a qual chamaram de “seleção do limiar pelo mínimo cruzamento da entropia”. Este método difere dos outros apresentados anteriormente por usar um conhecimento sobre o problema. A entropia cruzada (14) entre a distribuição de probabilidade a priori (conhecimento)  $p_x$  e a distribuição de probabilidade a posteriori  $q_x$  é que será usada na determinação do limiar.

$$H_{CE}(q, p) = \sum_{x=1}^G q_x \log \frac{q_x}{p_x} \quad (14)$$

$G$  é o número de níveis de intensidade e  $H_{CE}$  é chamada de entropia cruzada, entropia relativa, número de Kullback-Leibler, discriminação de informação, ou ainda divergência direta. No contexto de segmentação de imagens  $p_x$  representa algum conhecimento sobre a resposta correta e  $q_x$  é a informação que será processada, isto é, um determinado nível de intensidade.

O objetivo deste método é encontrar, para qualquer diferença entre  $p_x$  e  $q_x$ , uma solução mais próxima do resultado desejado ( $p_x$ ), pela minimização da diferença (entropia cruzada) dos elementos envolvidos. Aplicado a limiarização de imagens temos a seguinte equação:

$$H_{CE}(T) = \left[ \sum_{i=1}^G \mu_0(T) \log \frac{\mu_0(T)}{g_i} \right]_{g_i \leq T} + \left[ \sum_{i=1}^G \mu_1(T) \log \frac{\mu_1(T)}{g_i} \right]_{g_i > T} \quad (15)$$

Sendo:

$$\mu_0(T) = \sum_{i=1}^T i p_i \quad \mu_1(T) = \sum_{i=T+1}^G i p_i \quad p_i = \frac{g_i}{N}$$

$$\sum_{i=1}^G g_i = \left[ \sum_{i=1}^G \mu_0(T) \right]_{g_i \leq T} + \left[ \sum_{i=1}^G \mu_1(T) \right]_{g_i > T}$$

$g_i$  representa o número de pixels com intensidade  $i$ .

$G$  representa o número de intensidades de cinza da imagem.

$p_i$  representa a probabilidade da intensidade  $i$  ser encontrada na imagem.

$N$  representa a quantidade total de pixels da imagem.

Dessa forma, todo o espaço de busca é percorrido com o objetivo de encontrar o nível de intensidade  $T$  que minimize a equação (16), sendo este o limiar escolhido pelo modelo.

$$\tau_{CE} = \arg(\min\{H_{CE}(T)\}) \quad (16)$$

No mesmo ano, Brink [BRINK, 1996] propõe uma variação do trabalho desenvolvido por Pun e Kapur, usando novamente a entropia relativa, porém usando como conhecimento a variância dos níveis de intensidade da vizinhança do pixel e assim incorporando a esta abordagem a informação espacial na escolha automática do limiar.

(i-1, j-1)	(i, j-1)	(i+1, j-1)
(i-1, j)	(i, j) origem	(i+1, j)
(i-1, j+1)	(i, j+1)	(i+1, j+1)

Figura 7 - Vizinhança 3x3 de um pixel na coordenada (i,j) da imagem

A variância (17) é determinada na 8-vizinhança em relação ao pixel da origem, como mostra a Figura 7.



$$\sigma_{ij}^2 = \frac{\sum_{x=i-1}^{i+1} \sum_{y=j-1}^{j+1} (g_{xy} - \mu_{v8})^2}{9} \quad (17)$$

Sendo  $\sigma_{ij}^2$  a variância encontrada na 8-vizinhança,  $\mu_{v8}$  a média dos pixels da 8-vizinhança e  $g_{xy}$  a intensidade de cinza do pixel  $(x, y)$ , a nova entropia considerando a variância passa a ser obtida pela equação:

$$H = -\sum_{i=1}^L \sum_{j=1}^C g_{ij} \log \frac{g_{ij}}{m_{ij}}, \quad m_{ij} = 1 + \sigma_{ij}^2 \quad (18)$$

A variância  $m_{ij}$  é definida como  $1 + \sigma^2$ , para garantir que ela não será nula em todas as 8-vizinhanças. Esta informação é adicionada ao cálculo das entropias do objeto e fundo (19) da imagem.

$$\begin{aligned} H_b(T) &= -\sum_{i=1}^L \sum_{j=1}^C \frac{g_{ij}}{P_0(T)} \log \frac{g_{ij}/P_0(T)}{m_{ij}}, \\ H_w(T) &= -\sum_{i=1}^L \sum_{j=1}^C \frac{g_{ij}}{P_1(T)} \log \frac{g_{ij}/P_1(T)}{m_{ij}} \end{aligned} \quad (19)$$

Sendo que:

$$P_0(T) = -\sum_{i=1}^T i g_i \quad P_1(T) = -\sum_{i=T+1}^G i g_i$$

$g_{ij}$  representa a intensidade de cinza presente nas coordenadas  $(i, j)$  da imagem.

$g_i$  representa o número de pixels com intensidade  $i$ .

$L$  representa o número de linhas da imagem.

$C$  representa o número de colunas da imagem.

$G$  representa o número de intensidades de cinza da imagem.

$H$  representa a entropia relativa da imagem.

Encontradas as duas entropias  $H_b(T)$  e  $H_w(T)$ , o limiar é determinado pelo procedimento de *maxmin* que, segundo Brink [BRINK, 1996], é o procedimento que

divide igualmente as entropias entre objeto e fundo da imagem. Assim a intensidade de cinza  $T$  que “*maxminimiza*” a função  $\psi(T)$  (20) será o limiar escolhido, mas agora considerando uma informação espacial.

$$\psi(T) = MAX ( MIN \{ H_b(S), H_w(S) \} ) \quad (20)$$

Este incremento proporciona ao modelo a melhoria da informação visual contida na imagem limiarizada, pois segundo [BRINK, 1996] a inclusão da variância  $m_i$  representa um fator de ponderação que proporciona a preservação de informações visuais importantes como bordas, fundamentais para a compreensão visual da imagem.

Cada um dos métodos apresentados nesta seção tem suas vantagens e desvantagens em relação a algum critério de avaliação. Assim alguns comentários são válidos.

O método proposto por Pun, tem seu pior funcionamento com histogramas bimodais, se consideradas a uniformidade e forma da imagem limiarizada e também não produz bons resultados em comparação com outros métodos [SAHOO et al., 1988].

O método proposto por Johannsen e Bile apresenta bons resultados com histogramas bimodais e bimodais discretos. Seu desempenho pode ser comparado ao método de Otsu (seção 2.2.1). Essas considerações foram baseadas nos critérios de uniformidade e forma [SAHOO et al., 1988].

O método proposto por Kapur tem boas respostas para histogramas bimodais discretos. Para histogramas não bimodais apresenta bom desempenho no quesito forma, sendo classificado por Sahoo [SAHOO et al., 1988] como um método muito consistente para esta classe de imagens. O trabalho de Lee [LEE et. al., 1990] complementa observando que o desempenho deste método é dependente das distribuições do objeto e fundo da imagem. Em seu trabalho Kapur [KAPUR et. al., 1985] observa que a suavização (borramento) da imagem de entrada não afeta o desempenho de seu método, mas por outro lado Lee [LEE et. al., 1990] observa que é muito afetado (sensível) pela presença de ruído.

Um ponto positivo no método proposto por Beghdadi [BEGHDADI et. al., 1995] é que ele explora a correlação espacial dos pixels dentro de cada bloco, contudo entre os blocos ela não é considerada. Um grande problema que deve ser observado é a escolha do tamanho dos blocos, que influencia diretamente o desempenho do método.

O acréscimo da variância no cálculo da entropia proposto por [BRINK, 1996] representa sem dúvida um ganho adicional de informação na busca pelo limiar. Com base na qualidade visual, as respostas deste modelo se mostraram melhores que o método de Kapur, mas com aumento do custo computacional.

#### **2.2.4. Métodos Entrópicos Bidimensionais**

O precursor deste tipo de método foi Abutaleb [ABUTALEB, 1989] que em seu trabalho, propôs um método que pode ser visto como uma continuação dos métodos propostos por Pun [PUN, 1980] e Kapur [KAPUR et. al., 1985]. Sua contribuição está em acrescentar ao histograma alguma informação espacial da imagem.

Nestes métodos acrescenta-se uma dimensão ao histograma, tal como o agrupamento dos pixels que possuem a mesma média de intensidade em suas vizinhanças, gerando assim, um histograma bidimensional (2D).

Um histograma 2D adequado forma uma superfície com dois picos (Figura 8), um correspondendo ao objeto e o outro ao fundo. Os elementos próximos à diagonal principal da superfície contribuem mais para as representações do objeto e do fundo da imagem, enquanto a informação fora dessa diagonal representa melhor as bordas e ruídos.

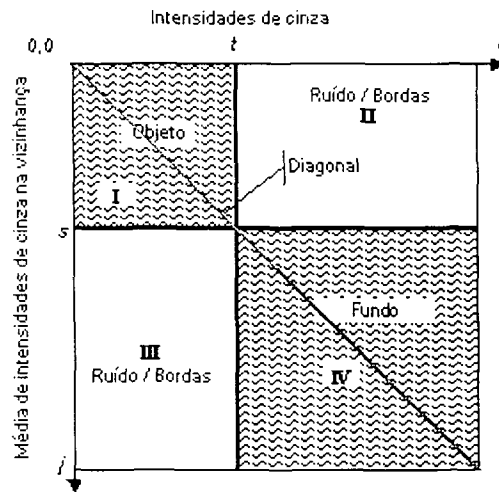


Figura 8 - Superfície gerada pelo histograma bidimensional.

A partir do histograma bidimensional é realizada uma busca exaustiva, com o objetivo de encontrar a média de vizinhanças ( $S$ ) e a intensidade de cinza ( $T$ ) que maximiza a função de separação entre o objeto e fundo da imagem (21), considerando a entropia contida no histograma 2D.

$$\psi(S, T) = H_o(S, T) + H_b(S, T) \quad (21)$$

Sendo:

$$H_b(S, T) = - \sum_{i=1}^S \sum_{j=1}^T (p_{ij} / P_{ST}) \log(p_{ij} / P_{ST}) \quad H_w(S, T) = - \sum_{i=S+1}^G \sum_{j=T+1}^G (p_{ij} / (1 - P_{ST})) \log(p_{ij} / (1 - P_{ST}))$$

$$P_{ST} = - \sum_{i=1}^S \sum_{j=1}^T p_{ij} \quad p_{ij} = \frac{g_{ij}}{N^2}$$

$g_{ij}$  representa o número de pixels com média  $i$  e intensidade  $j$ .

$G$  representa o número de intensidades de cinza da imagem.

$p_{ij}$  representa a probabilidade da média  $i$  e a intensidade  $j$  serem encontrados juntos na imagem.

$N$  representa o número total de pixels da imagem.

Brink [BRINK, 1992] em seu trabalho, analisou a forma de escolha do limiar proposta por Abutaleb e constatou que ele simplesmente maximiza a soma das entropias. Ele propôs então que fosse usada a busca do “*maxmin*”, para garantir que a

divisão fosse equitativa entre as entropias do objeto e do fundo da imagem. O limiar é determinado então pelos valores  $(S, T)$  que “*maximizem*”, conforme (22).

$$\psi(s, t) = \text{MAX}(\text{MIN}\{H(A), H(B)\}) \quad (22)$$

Embora a inclusão de uma nova característica na escolha do limiar realmente melhore o desempenho da limiarização, também aumenta o tempo computacional necessário para determiná-lo [ABUTALEB, 1989], considerando que os dois métodos apresentados têm suas implementações sequenciais.

Outra consideração importante é que o método bidimensional se mostrou mais robusto que o unidimensional, com relação à presença de ruídos na imagem [ABUTALEB, 1989]. Entre os dois métodos bidimensionais, foi observado que, baseada na avaliação subjetiva de reconhecimento das imagens limiarizadas, a abordagem usando “*maxmin*” em geral tem melhor desempenho. Por outro lado, constatou-se que, para imagens de infravermelho, a abordagem de Abutaleb pode produzir melhores resultados [BRINK, 1992].

### 2.2.5. Método Transformação do Histograma

Este método é descrito em [SAHOO et al., 1988] e tem uma particularidade interessante: não seleciona diretamente o limiar para a imagem, mas transforma o histograma. Esta transformação consiste em tornar os vales mais profundos e os picos mais altos e, dessa forma, permitir a aplicação do método *Mode* (apresentado na seção 2.2) na escolha do limiar.

Uma técnica comumente usada neste método é a ponderação dos pixels da imagem de acordo com as propriedades locais, como por exemplo, o uso de um operador de borda [MASON et. al., 1975] que atribua aos pixels contidos em regiões homogêneas mais peso do que aqueles que estiverem em regiões de bordas.

Outro exemplo é o uso do método *quadtree* [WU & ROSENFELD, 1982], o qual divide a imagem original em quadrantes e então mede o desvio padrão das intensidades de cinza contidas em cada um deles. Se a região for homogênea, o desvio

padrão é baixo, caso contrário é alto. Assim, cada quadrante com desvio padrão alto é novamente dividido até que o desvio padrão não ultrapasse um valor previamente escolhido. Após a decomposição, um novo histograma é obtido com as médias de cada bloco.

O objetivo das técnicas de transformação é produzir, a partir do histograma original da imagem, um novo histograma com vales e picos realçados. Para isso considera alguma propriedade local dos pixels. Isto o torna mais robusto, pois passa a considerar a disposição espacial da informação.

Uma observação importante é que pode ser aplicado a imagens que possuem histograma unimodal, as quais normalmente são de difícil limiarização; por outro lado apenas gera um novo histograma ponderado, o qual será usado por outro método de limiarização [SAHOO et. al., 1988]. Sua implementação é sequencial e após a transformação do histograma pode ser usado tanto por alguma técnica de limiarização global quanto por alguma técnica de limiarização local, como por exemplo, o método *Mode* (apresentado na seção 2.2).

#### **2.2.6. Métodos de Segunda Ordem dos Níveis de Intensidade**

Os métodos de segunda ordem dos níveis de intensidade recebem esse nome por considerarem a informação não somente do histograma da imagem (estatística de primeira ordem dos níveis de intensidades), mas também outras informações estatísticas, como a matriz de co-ocorrência, introduzida por Haralick [HARALICK et al., 1973] para a análise de textura.

Na matriz de co-ocorrência, os seus elementos são dispostos de acordo com a vizinhança e com o número de ocorrências [GONZALEZ & WOODS, 2000]. Esta análise é possível, pelo fato da matriz de co-ocorrência  $M_{(d,\phi)}$  ser formada pelo agrupamento de pixels de mesma intensidade (ordem da matriz), organizados por uma distância  $d$  e sentido (orientação)  $\phi$  da ocorrência das demais intensidades presentes na imagem.

Sendo a formação da matriz de co-ocorrência dependente de  $(d, \phi)$ , a presença de uma determinada textura pode ser detectada pela escolha de valores apropriados para distância e ângulo. A matriz de co-ocorrência foi usada na seleção do limiar pelo trabalho de Ahuja e Rosenfeld [AHUJA & ROSENFELD, 1978], o qual define a matriz como:

$$M_{(d, \phi)} = M_{(1, 0)} + M_{(1, \pi/2)} + M_{(1, \pi)} + M_{(1, 3\pi/2)} \quad (23)$$

Esta definição identifica que uma determinada intensidade  $i$  aparece como um 4-vizinho do pixel com intensidade  $j$ . Dessa forma, os pixels que formam o objeto (região homogênea), influenciarão mais os elementos da diagonal da matriz, pois terão 4 vizinhos com intensidades semelhantes (homogeneidade do objeto ou fundo), e os pixels de bordas deverão influenciar mais os elementos fora da diagonal da matriz, por haver mais mudanças de intensidades. Assim a matriz de co-ocorrência pode ser usada na definição de dois novos histogramas [SAHOO et. al., 1988]:

1. Histograma baseado nos elementos da proximidade da diagonal da matriz, o qual deverá ter um vale profundo entre as intensidades do objeto e do fundo.
2. Histograma baseado nos elementos fora da diagonal da matriz, o qual deverá ter um pico agudo entre o objeto e o fundo da imagem.

Um limiar para a imagem pode então ser escolhido na intensidade mais profunda do vale na opção 1 ou na intensidade mais alta do pico no caso 2.

Uma alternativa semelhante à matriz de co-ocorrência foi desenvolvida por Kirby e Rosenfeld [KIRBY & ROSENFELD, 1979], que ao invés de considerar a distância e sentido entre os pixels da imagem, considera a intensidade do pixel e sua média local na formação de uma matriz (médias x intensidades). Inicialmente com a maior média e menor intensidade, e depois decrementando-as (linhas) e incrementando-as (colunas) respectivamente.

Dessa forma, os elementos próximos à diagonal da matriz representam os pixels com intensidade e média próximos, ou seja, representam os pixels que formam o objeto ou o fundo da imagem pela sua homogeneidade. O histograma gerado pelos elementos próximos da diagonal da matriz, deverá ter um vale separando o objeto do fundo da imagem.

Por outro lado, os elementos distantes da diagonal da matriz são aqueles com intensidade e média local distantes, o que significa que estão próximos das bordas da imagem. A composição de um histograma com esses elementos normalmente forma um pico agudo separando o objeto do fundo da imagem.

Uma vez formado o histograma pelos elementos próximos da diagonal, o limiar pode ser escolhido no intervalo entre as intensidades do vale ou, se formado pelos elementos distantes da diagonal, o limiar pode ser escolhido entre as intensidades que formam o pico [SAHOO et. al., 1988].

Os métodos apresentados nesta seção apresentam implementação sequencial, exigem muito tempo computacional e não são indicados para aplicações que envolvam tempo real. Por outro lado, são capazes de inserir um pré-conhecimento do que se busca, e assim, ter melhor desempenho para um conjunto de imagens com padrão conhecido e particularmente o método desenvolvido por Kirby e Rosenfeld é capaz de transformar um histograma unimodal em bimodal simplesmente acrescentando a média das intensidades.

Imagens cujos histogramas não são bimodais, para este método, apresentam melhor desempenho no quesito forma, mas o desempenho é sofrível no quesito uniformidade. Isto classifica este método como sendo orientado à forma [SAHOO et al., 1988].

### **2.2.7. Métodos com Abordagens Evolucionárias**

Este método pode ser aplicado na busca de um limiar global ou multiníveis. O Algoritmo Genético (AG) representa um mecanismo de busca robusto, desde que o espaço em que se procura não seja contínuo. A possibilidade de implementação



paralela de AGs torna a busca por limiares muito mais rápida que os métodos iterativos (seqüenciais) [YIN, 1999].

Os operadores genéticos: mutação, cruzamento, tamanho da população e número de gerações são parâmetros deste modelo. A população inicial é aleatoriamente escolhida.

A função-objetivo representa um determinado critério. Assim, a função-objetivo pode generalizar qualquer método que se baseie na busca em um espaço de características, como é o caso do método de máxima entropia [KAPUR et al., 1985] ou da maximização das variâncias das intensidades dos pixels [OTSU, 1978], usados no trabalho de Yin [YIN, 1999].

Com o objetivo de acelerar o processo de busca do AG, Yin propõe uma estratégia de aprendizado. A estratégia consiste no armazenamento da melhor solução de cada geração e baseado neste indivíduo, criar outros dois: um pela soma (+1) e outro pela subtração (-1) de valores. Em seguida é verificado se o resultado é melhor e assim o melhor indivíduo (o mais bem adaptado) é armazenado e comparado com o melhor da próxima geração, e assim por diante. Este processo se repete até atingir alguma condição de parada típica de algoritmos evolucionários.

Se a abordagem exigir uma busca do melhor limiar, fica patente a eficiência computacional dos AGs, permitindo inclusive soluções implementadas sobre arquiteturas paralelas.

#### **2.2.8. Métodos com Lógica Nebulosa**

A teoria de conjuntos nebulosos incorpora a manipulação de incertezas, as quais são causadas pelas deficiências de informação como: informações incompletas, imprecisas, mal definidas, vagas ou contraditórias. Sua aplicação na limiarização de imagens tem a mesma motivação: as incertezas causadas pela ocorrência de bordas e regiões mal definidas da imagem, imagens de baixa qualidade e imagens afetadas por ruído [CHENG & CHEN, 1999].

No processo de limiarização de imagens com objetos bem definidos, como fundo claro e objeto(s) escuro(s), ou vice-versa, a determinação de um limiar que melhor separe as duas classes é uma tarefa relativamente fácil, mas isso dificilmente ocorre em imagens tomadas em ambientes desprovidos de controle.

A idéia essencial representada pelos conjuntos nebulosos está no relaxamento de decisões rígidas (ou rigorosas) que condiciona um grau de pertinência a um elemento (no caso, um pixel da imagem) a uma determinada classe. Isto permite tornar mais realística a interpretação do problema e tratar as ocorrências de elementos com níveis intermediários de pertinência [JAWAHAR et al., 2000]. O nível ou grau de pertinência em geral é representado por um número real no intervalo  $[0,1]$ .

Assim como na limiarização global, o objetivo é encontrar duas regiões distintas da imagem:  $O$  e  $F$ , correspondendo a classes objeto e fundo, porém na limiarização nebulosa considera-se o grau de pertinência dos pixels, como mostra a equação (24).

$$O = \{ \mu_O(x_i) / x_i, \quad i = 1, 2, \dots, N \} \quad \text{e} \quad F = \{ \mu_F(x_i) / x_i, \quad i = 1, 2, \dots, N \} \quad (24)$$

A função  $\mu_O(x_i)$  representa o nível de pertinência do elemento  $x_i$  em relação ao objeto e  $\mu_F(x_i)$  o nível de pertinência do elemento  $x_i$  em relação ao fundo.  $N$  é o número de níveis de intensidade. A decisão de classificação é resultante do maior valor entre as respectivas funções de pertinência.

$$I = \begin{cases} 0 & \text{se } \mu_O(x_i) > \mu_F(x_i) \\ 255 & \text{se } \mu_F(x_i) \leq \mu_O(x_i) \end{cases} \quad i = 1, 2, \dots, N \quad (25)$$

Na teoria de conjuntos nebulosos, quando um pixel  $g$  é classificado como objeto, diz-se que  $g$  pertence ao objeto ou que  $g$  é membro da classe objeto.

Desta forma, as funções de pertinência  $\mu_O(x_i)$  e  $\mu_F(x_i)$  influenciam diretamente a limiarização das imagens. Existem várias funções de pertinência

[PEDRYCZ, 1998], como por exemplo, as funções  $S$  e  $Z$ . A função  $S$  é mostrada em (26).

$$S(x, a, b, c) = \begin{cases} 0, & x \leq a \\ \frac{(x-a)^2}{(b-a)(c-a)}, & a \leq x \leq b \\ 1 - \frac{(x-c)^2}{(c-b)(c-a)}, & b \leq x \leq c \\ 1, & c \leq x \end{cases} \quad (26)$$

Sendo  $x$  um nível de cinza entre 0 e 255, os parâmetros  $a$ ,  $b$  e  $c$  determinam a forma da função de pertinência  $S$ . O intervalo formado entre os parâmetros  $a$  e  $c$  é chamado região nebulosa. O parâmetro  $b$  é normalmente o ponto médio do intervalo  $[a, c]$ , mas não necessariamente. Quando isto ocorre, a função  $S$  é chamada de função  $S$  padrão. A função  $Z$  é obtida a partir da função  $S$ , como mostra (27).

$$Z(x, a, b, c) = 1 - S(x, a, b, c) \quad (27)$$

A atribuição de uma classe a um elemento depende de seu grau de pertinência a esta classe. Na Figura 9 ilustra-se a aplicação das funções de pertinência  $S$  e  $Z$  no histograma de uma imagem em níveis de cinza.

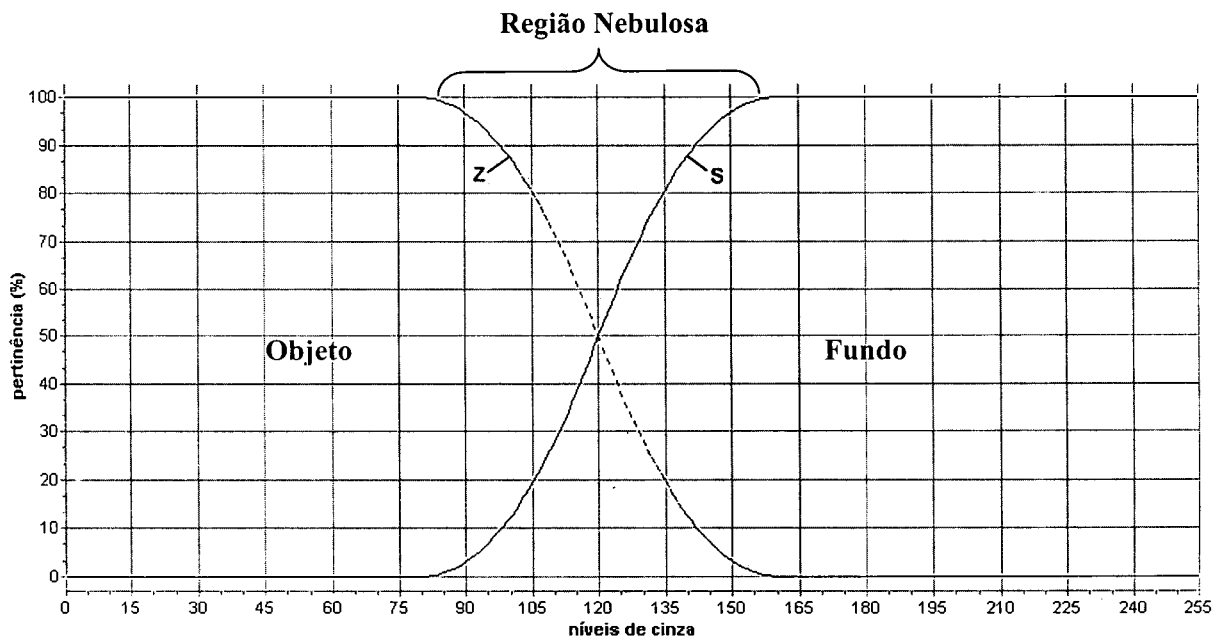


Figura 9 - Funções de pertinência nebulosas:  $S$  e  $Z$ .

Na Figura 9, os valores utilizados para os parâmetros  $a$ ,  $b$  e  $c$  foram (80,120,160). Pode ser vista a região nebulosa entre as pertinências das funções no intervalo  $[a, c] = [80, 160]$ .

Existem basicamente duas tarefas neste tipo de limiarização: a determinação da região nebulosa e a atribuição do grau de pertinência. Como pode ser observado na Figura 9, a determinação da região nebulosa é uma determinante neste modelo de limiarização e como pode-se notar, a interseção entre os conjuntos não é vazia. Portanto, a limiarização nebulosa generaliza a limiarização global, que se torna um caso especial quando a interseção entre objeto e o fundo é vazia.

Na região nebulosa, o grau de pertinência é atribuído aos elementos cuja classificação entre objeto e fundo é incerta. A localização do ponto  $b$  influencia diretamente a pertinência para as funções  $\mu_O(x)$  e  $\mu_F(x)$ . Este método de limiarização fica, dessa forma, totalmente dependente da combinação dos valores dos parâmetros  $a$ ,  $b$  e  $c$ .

A função de pertinência pode ser determinada por retas definidas por equações do tipo mostrado em (28).

$$y = ax + b \quad (28)$$

A Figura 10 exemplifica a aplicação desta função de pertinência.

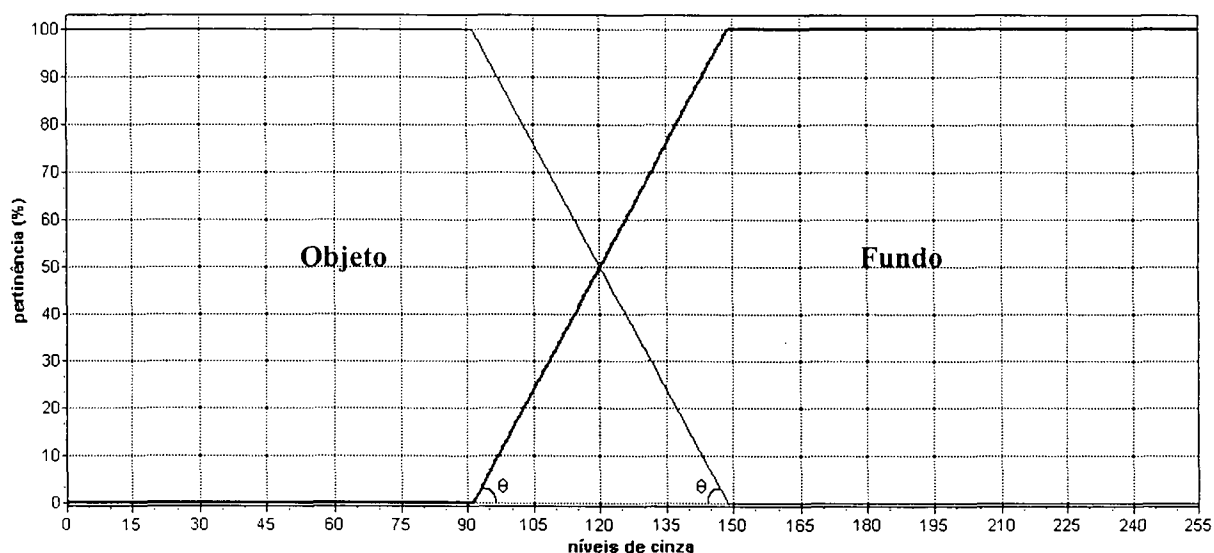


Figura 10- Funções de pertinência por retas.

Como se pode observar na Figura 10, os ângulos de inclinação das retas são os mesmos ( $60^\circ$ ), garantindo que a soma das pertinências seja sempre 100%. Os valores usados na Figura 10 foram:  $y = 50$ ,  $x = 120$ , o que resultou na região nebulosa entre as intensidades 91 e 149.

O grau de pertinência é então uma função dos níveis de intensidade que pode ser parametrizada pelo ângulo  $\theta$  de inclinação das retas e usar a equação reduzida da reta para encontrar os coeficientes angular e linear.

$$\begin{aligned}
 y &= ax + b && (\text{equação reduzida da reta}) \\
 b &= y - ax \\
 b &= 50 - a \times 120 \\
 b &= 50 - \tan(\theta) \times 120 \\
 b &= 50 - \tan(60^\circ) \times 120
 \end{aligned} \tag{29}$$

Uma vez determinados os valores para os coeficientes angular ( $a$ ) e linear ( $b$ ), é possível encontrar o início e o fim da região nebulosa, ou seja, o valor em que as retas cortam o eixo dos níveis de intensidade.

Usando a teoria dos conjuntos nebulosos, Fleury [FLEURY et. al., 1996] propõe uma abordagem na determinação de limiares multiníveis. Seu método divide-se entre a procura por picos, a definição das regiões nebulosas e a busca pela mínima entropia de cada região.

Antes de localizar os picos é proposta a suavização do histograma, como uma forma de eliminar a presença de falsos picos e vales. Em seguida é iniciada a busca por picos no histograma pela técnica de recorte (*histogram clipping*) [GONZALEZ & WOODS, 2000], e então os locais dos picos são armazenados. Em seguida, o histograma é dividido em várias regiões, cada uma delas delimitadas pelos picos encontrados, ou seja,  $região_n = [pico_n, pico_{n+1}]$ .

Em cada região é aplicada uma função S-padrão, sendo os valores dos parâmetros  $a$  e  $c$  determinados pelos picos que formam a região e  $b = (a+c)/2$ . Assim a entropia nebulosa de cada região é encontrada pela equação (30).

$$H_{fuzzy} = \frac{1}{n \log(2)} \sum_{i=1}^n Sn(\mu(g_i)) h_i \quad Sn(x) = -\mu(x) \log_2 \mu(x) \quad (30)$$

Dessa forma, o limiar para cada região é determinado pelo nível de intensidade  $x$  que satisfaz a condição imposta pela equação (31), e então encontrando o vale da região.

$$Sn(x-1) > Sn(x) < Sn(x+1) \quad (31)$$

No trabalho de Fleury também é proposta a implementação paralela do algoritmo, o que o torna computacionalmente bastante eficiente. É aplicado em imagens com histograma bimodal e multimodal, com resultados, em geral, melhores que os resultados por máxima entropia [KAPUR et. al., 1985] para imagens multimodais. Outra observação importante é que a abordagem nebulosa se mostrou mais robusta que o método de Kapur [FLEURY et. al., 1996].

### 2.2.9. Método Aprendizado por Reforço

O método de limiarização de imagens usando aprendizado por reforço foi proposto por Yin [YIN, 2002], implementando um modelo básico de aprendizagem por reforço (BRLM, *basic reinforcement learning model*) e usando o algoritmo *Q-Learning* [WATKINS & DAYAN, 1992; KAEHLBLING et. al., 1996].

Esta abordagem usa um agente simples e uma tabela contendo um conjunto de estados. Para cada estado são armazenadas as possíveis ações que podem ser tomadas e o reforço para cada ação (positivo ou negativo). O agente então pode aprender uma tarefa (escolha do limiar) por uma sequência de tentativas e erros/acertos no ambiente, isto é, no histograma.

O aprendizado ocorre da seguinte forma: o agente percebe o estado corrente do ambiente e então escolhe qual ação executar baseado no maior reforço positivo da tabela para aquele estado. O estado do ambiente é alterado pela ação do agente, originando um novo estado. Então o agente recebe um reforço pela transição do estado. Se o novo estado for melhor, o reforço é positivo, caso contrário é negativo.

A cada transição de estados, a tabela é atualizada de acordo com o reforço recebido. Este processo de aprendizagem é repetido até que uma condição de parada seja alcançada, ou seja, se o agente atingiu o objetivo ou chegou a um estado terminal. A Figura 11 apresenta uma representação gráfica do funcionamento deste processo.



Figura 11 - Modelo básico de aprendizado por reforço

Em seu trabalho, Yin aplica uma aprendizagem BRLM para determinar tanto o limiar global quanto o multiníveis. No caso do limiar global, o histograma é definido como ambiente e o agente recebe um estado inicial, por exemplo, ficando posicionado na intensidade 0. Em seguida, ele executa uma ação, por exemplo, caminhar até a intensidade 150. Assim o reforço será dado de acordo com o cálculo da entropia da posição anterior (intensidade = 0) até a posição atual (intensidade = 150) e este reforço é atualizado na tabela que faz o papel de memória do agente.

O agente é então estimulado a encontrar o estado e a ação que maximiza o reforço, que neste caso será a entropia. Neste ponto fica claro que é empregado o BRLM como uma estratégia de busca pela máxima entropia [PUN, 1980; KAPUR, 1985] (seção 2.2.3) ao invés da busca por força bruta proposta inicialmente pelos criadores do método.

Este mecanismo de busca também pode ser estendido para os métodos entrópicos bidimensionais [ABUTALEB, 1989] (seção 2.2.4) e, neste caso, o agente deverá executar duas ações para mudar de estado ao invés de apenas uma, i.e., escolher o nível de intensidade e a média das intensidades.

Segundo Yin [YIN, 2002], a estratégia de busca usando BRLM se mostrou mais eficiente que a abordagem usando algoritmo genético desenvolvida por ele mesmo [YIN 1999].

## 2.3 Algoritmos Básicos

Os atributos apresentados na seção anterior dependem de algoritmos para serem extraídos das imagens. Nesta seção são apresentados resumidamente alguns dos algoritmos utilizados neste trabalho.

### 2.3.1. Histograma de Níveis de Intensidade

O histograma de uma imagem digital é a distribuição das intensidades, ou seja, o número de pixels para cada um dos níveis de intensidade presentes na imagem. A Figura 12 mostra um exemplo de uma imagem e seu respectivo histograma.

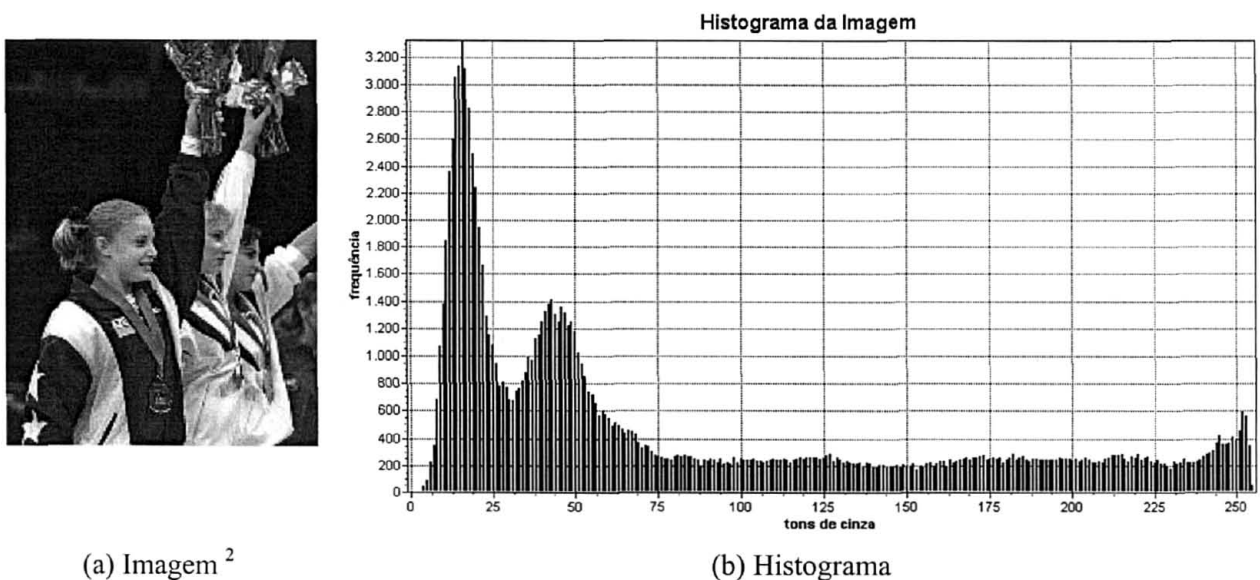


Figura 12 - Imagem e seu Histograma.

---

<sup>2</sup> PRINCETON. Department of Computer Science - Princeton University [http://www.cs.princeton.edu/~dbradley/profile\\_images1.zip](http://www.cs.princeton.edu/~dbradley/profile_images1.zip)



Um algoritmo típico, em pseudo-código, para o cálculo do histograma de uma imagem digital é apresentado no Algoritmo 1.

<pre> <u>Tipo</u> histograma = vetor[0..255] de inteiros;  <u>Função</u> ghistograma (Img : imagem) : histograma; <u>Variáveis</u>   x,y, cinza : inteiro;   h : histograma; <u>Início</u>   <u>Para</u> y = 1 <u>Até</u> Img.altura <u>Faça</u>     <u>Para</u> x = 1 <u>Até</u> Img.largura <u>Faça</u>       cinza = Img(x,y);       h[cinza] = h[cinza]+1;     <u>Fim-Para</u>   <u>Fim-Para</u>   <u>Retorne</u>(h); <u>Fim.</u> </pre>
Algoritmo 1 - Histograma de uma imagem.

### 2.3.2. Equalização do Histograma

A equalização é aplicada ao histograma da imagem com o objetivo de tornar a distribuição de intensidades de cinza mais uniforme ao longo do histograma. A consequência dessa uniformidade é o aumento de contraste.

A equalização é obtida mapeando-se cada nível de intensidade presente na imagem original em um outro nível. Esse mapeamento é resultante da aplicação da função de distribuição acumulada (FDA) na frequência de ocorrência de um determinado nível. A FDA é obtida acumulando-se as frequências de ocorrência dos níveis de intensidade menores que um determinado nível  $i$  e dividindo-se este total acumulado pelo número total de pixels na imagem.

$$Ce_k = \sum_{i=0}^k \frac{f(i)}{N} \quad (32)$$

Sendo:

$Ce_k$  representa a intensidade  $k$  equalizada pela FDA.

$f(i)$  representa o número de ocorrências da intensidade  $i$  na imagem original.

$N$  representa o número de pixels da imagem.

Na verdade, a equalização é uma tentativa de transformar a distribuição das intensidades dos pixels em uma distribuição uniforme.

O Algoritmo 2 apresenta a implementação da equalização de uma imagem. Um exemplo da aplicação desse algoritmo pode ser visto na Figura 14, a qual apresenta a imagem equalizada (a), a FDA (b) e o histograma equalizado (c). Esse exemplo teve como entrada as informações apresentadas pela Figura 13.

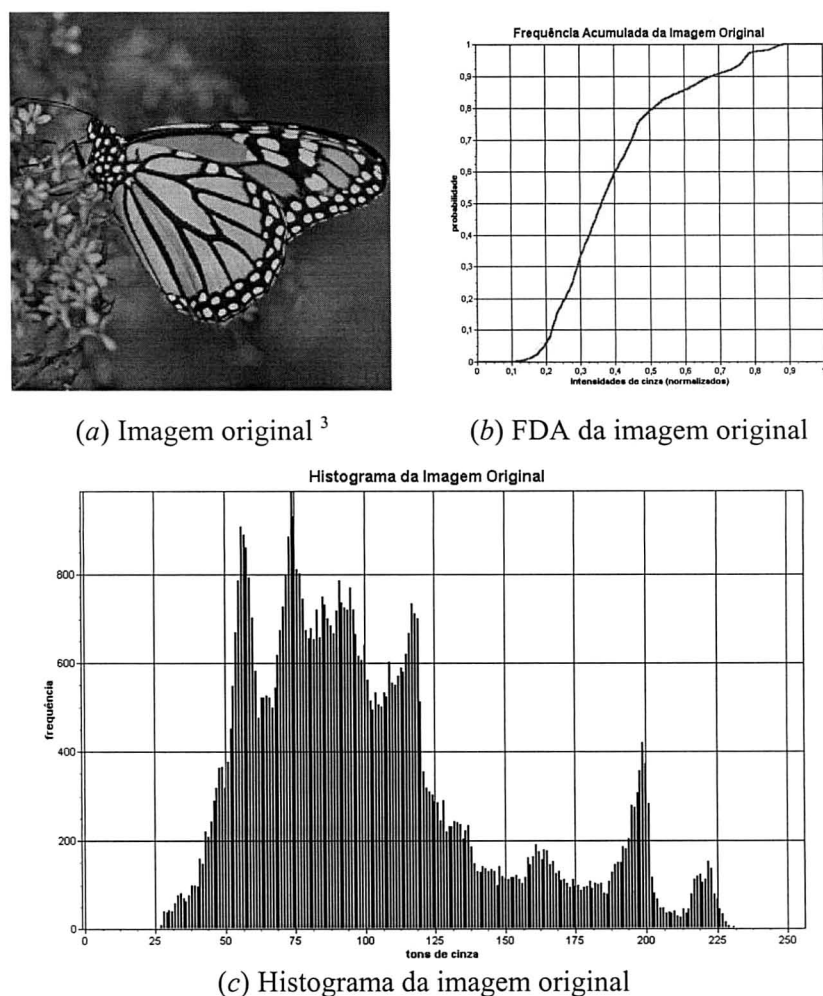


Figura 13 - Imagem original, FDA e seu Histograma.

A equalização é normalmente utilizada por métodos automáticos de limiarização, como os que são apresentados na seção 2.2.3. Nestes métodos a equalização impede a ocorrência de platôs ou mínimos/máximos locais nas funções que definem o critério de limiarização. O efeito da alteração no contraste de uma imagem equalizada pode ser observado na Figura 13 e Figura 14.

<sup>3</sup> SAMPL. Signal and Machine Perception Laboratory – Departament of Eletrical Engineering – The Ohio State University. <http://sAMPL.eng.ohio-state.edu/~sAMPL/database.htm>

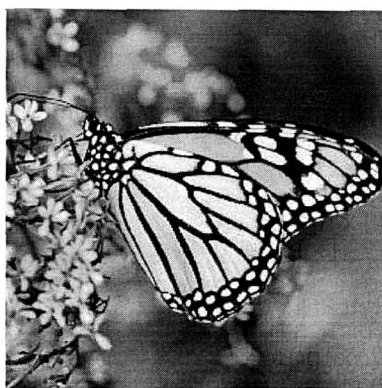
```

Tipo histograma = vetor[0..255] de inteiros;
Função equaliza (Img : imagem; h : histograma) : imagem;
variáveis  ImgE : imagem;
           N, x,y, cinza : inteiro;
           freqacumulada, he : histograma;

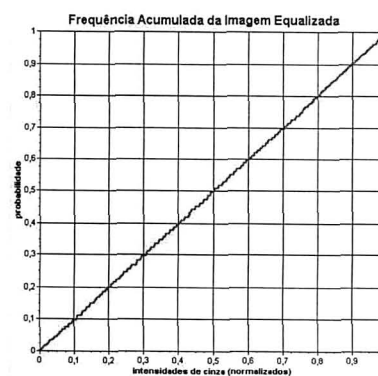
Início
  N = Img.altura * Img.largura;
  Para x = 0 até 255 faça
    acum = acum + h[x] / N;
    freqacumulada[x] = acum;
  Fim-Para
  Para y = 1 até Img.altura faça
    Para x = 1 até Img.largura faça
      cinza = Img(x,y);
      ImgE(x,y) = freqacumulada[cinza] * 255;
    Fim-Para
  Fim-Para
  Retorne(ImgE);
Fim.

```

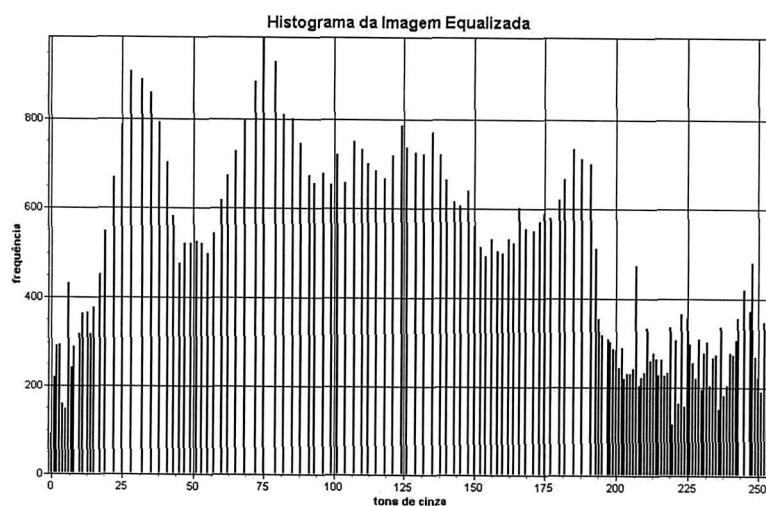
Algoritmo 2 - Equalização de uma imagem.



(a) Imagem equalizada <sup>4</sup>



(b) FDA da imagem equalizada



(c) Histograma da imagem equalizada

Figura 14 - Imagem equalizada, FDA e seu Histograma.

<sup>4</sup> SAMPL. Signal and Machine Perception Laboratory – Department of Electrical Engineering – The Ohio State University. <http://sAMPL.eng.ohio-state.edu/~sAMPL/database.htm>

### 2.3.3. Exemplo de Limiarização

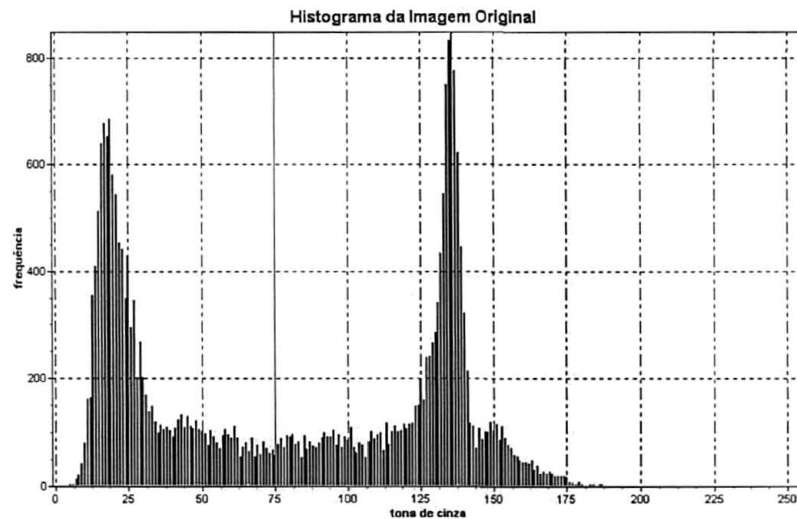
A limiarização de imagens pode ser global ou multiníveis. Nesta seção será apresentado o algoritmo de limiarização global. A limiarização multiníveis pode ser vista como uma extensão do limiar global.



(a) Imagem em intensidades de cinza <sup>5</sup>



(b) Imagem limiarizada na intensidade  $T=75$ .



(c) Histograma da Imagem (a)

Figura 15 - Limiarização de Imagens

O resultado de uma limiarização global é a intensidade que classificará cada pixel como pertencente a um fundo ou a um objeto. Um algoritmo típico de limiarização global é implementado no Algoritmo 3. Tomando como entrada a

---

<sup>5</sup> LIS. Laboratories of Intelligence Systems of Institute of Information Science at Academia Sinica, the national academic research institute in Taiwan. <http://smart.iis.sinica.edu.tw/html/face.html>

imagem da Figura 15.a e o limiar  $T = 75$ , a função implementada terá como saída a imagem exibida na Figura 15.b.

<pre> <b>Função</b> Limiarização (Im : Imagem; T : inteiro): Imt : Imagem; <b>variáveis</b>   x,y : inteiro; <b>Início</b>   Para y = 1 até Im.altura <b>faça</b>     Para x = 1 até Im.largura <b>faça</b>       Se Im(x,y) &gt; T <b>então</b>         Imt(x,y) = 255 {branco}       <b>Senão</b>         Imt(x,y) = 0 {preto}       <b>Fim-Para</b>     <b>Fim-Para</b>   <b>Fim.</b> </pre>
<p>Algoritmo 3 - Limiarização global genérico.</p>

#### 2.3.4. Um exemplo de limiarização automática: Máxima Soma das Entropias

Alguns métodos de limiarização automáticos são resumidamente descritos na seção 2.2. Nesta subseção será tomado como exemplo a implementação do método proposto por Kapur [KAPUR et. al., 1985]. Esse método é aplicado ao histograma da imagem equalizada. O limiar é escolhido como a intensidade que maximiza a soma das entropias (seção 2.1.1) de duas partes distintas (objeto e fundo) da imagem.

Assim, para encontrar o limiar de máxima entropia de uma determinada imagem, o algoritmo testa todos os limiares possíveis, ou seja, todas as intensidades. Deste modo, classifica-se os pixels de uma imagem digital em dois tipos, de acordo com intensidade: o primeiro tipo com intensidade entre 0 e  $t$ , e o segundo tipo entre  $t+1$  e 255 (ou o nível máximo). Para cada uma destas partes, calcula-se a entropia. Inicialmente o valor de  $t$  é zero, sendo ele incrementado a cada iteração até que alcance o nível máximo (255).

A cada iteração o algoritmo guarda, na posição de um vetor, a soma das entropias da primeira parte (entre 0 e  $t$ ) e da segunda (entre  $t+1$  e 255). Ao final, o limiar cujo valor no vetor é máximo é selecionado. O Algoritmo 4 descreve a implementação do modelo de limiarização automático por máxima entropia proposto por Kapur.

```

Função entropia(x : real) e : real;
Início
    MAX, H0, H1, P0, P1 : real;
    Se (x > 0) então
        e := -(x * Log2(x));
    Senão
        e := 0;
    Fim.

Função MaxEntropia (histograma [0..255] de inteiros; totalpixels :
inteiro) Limiar : inteiro;
variáveis
    MAX, H0, H1, P0, P1 : real;
    t, i : integer;
Início
    MAX = 0;
    Para t = 0 até 255 faça
        P0 = 0;
        P1 = 0;
        Para i = 0 até t faça
            P0 = P0 + histograma[i] / totalpixels;
        Para i = t+1 até 255 faça
            P1 = P1 + histograma[i] / totalpixels;
        H0 = 0;
        H1 = 0;
        Para i = 0 até t faça
            H0 = H0 + entropia ((histograma[i] / totalpixels)/P0);
        Para i = t+1 até 255 faça
            H1 = H1 + entropia ((histograma[i] / totalpixels)/P1);
        Se (H0+H1 > MAX) então
            MAX = H0+H1;
            Limiar = t;
        Fim-então;
    Fim-para;
Fim.

```

Algoritmo 4 - Limiar por máxima entropia.

Uma vez determinado o limiar, é aplicada a limiarização conforme o Algoritmo 3.

### 2.3.5. Algoritmo de Agrupamento K-médias

K-médias [CONSULARO, 2000] é um algoritmo classificador que divide as amostras em  $k$ -grupos. É um algoritmo particional e, portanto, não estabelece uma hierarquia estruturada dos grupos. Realiza um aprendizado não-supervisionado, isto é, na fase de treinamento as amostras não são comparadas com alguma classificação correta, mas determinadas por uma métrica de similaridade entre a amostra e os  $k$  grupos, sendo cada grupo representado por uma média dos seus elementos. Tipicamente, a métrica utilizada é a distância euclidiana entre a amostra e a média (ou centro), contudo outras métricas podem ser utilizadas.

A classificação pelo algoritmo  $k$ -médias se inicia com uma escolha aleatória de  $k$  amostras de todo o conjunto de amostras a serem classificadas. Assumem-se então estes  $k$  elementos como centros iniciais. Em seguida é calculada a distância de todas as amostras restantes do conjunto à cada um dos  $k$  centros. Cada amostra será agrupada ao conjunto de dados cujo centro  $C_i$  estiver mais próximo. Uma vez tendo todas as amostras pertencendo a um dos  $k$  grupos, um novo centro  $C_i$  é estabelecido recalculando-se a média dos elementos do grupo  $i$  (com  $i$  entre 1 e  $k$ ). Deste modo, cada um dos centros se desloca. Esse processo é repetido até que os centros não se desloquem mais, ou então, se desloquem muito pouco relativamente à iteração anterior, conforme descreve o Algoritmo 5.

```

tipo
  registroe = registro
    x,y : real;
    agrupado : booleano;
    centroagrupado : inteiro;
  fim-registro;

Função kmedias (k, nr_amostras:inteiro; amostras : [1..nr_amostras] de
  registroe):centros[1..k] de inteiros;
variáveis
  menor_distancia, aux : real;
  i, j, agrupado, a_agrupar, centro : integer;
Início
  i = 0;
  Enquanto (i < k) faça
    agrupado = sorteio (nr_amostras);
    Se (amostras[agrupado].agrupado = falso) então
      i = i + 1;
      centros[i] = amostras[agrupado];
    Fim-então;
  Fim-enquanto;
  a_agrupar = nr_amostras - k;
  Enquanto (a_agrupar <> 0) faça
    menor_distancia = DISTANCIAMAXIMA(centros, amostras);
    Para i = 1 até k faça
      Para j = 1 até nr_amostras faça
        Se (amostras[j] ainda não agrupada) então
          aux = DISTANCIAEUCLIDIANA(centros[i], amostras[j]);
          Se (aux < menor_distancia) então
            menor_distancia = aux;
            centro = i;
            agrupado = j;
          Fim-se;
        Fim-se;
      Fim-para;
    Fim-para;
    a_agrupar = a_agrupar-1;
    centros[centro] = CENTROIDE(centro[centro], amostras[agrupado]);
  Fim-enquanto;
Fim.

```

Algoritmo 5 - Algoritmo de agrupamento K-médias.

Neste trabalho, o algoritmo k-médias é aplicado na determinação dos centros usados pela rede de funções de base radial apresentado nas seções 2.4 e 3.3.2. O k-médias é um algoritmo não supervisionado que segundo Haykin [HAYKIN, 2001], determina os centros das funções de base radial nas regiões onde estão os dados significativos, mesmo quando aplicado a um conjunto de dados sem um comportamento pré-determinado, como é o caso da percepção visual humana aplicada na tarefa de limiarizar imagens digitais.

## 2.4 Funções de Base Radial (FBR) e Redes de FBRs

Uma função de base radial (FBR),  $\phi$ , recebe este nome por ser radialmente simétrica com relação a um centro  $\mu_c$ . Desse modo, ela pode ser definida como:  $\phi(r) = \phi(\|r - \mu_c\|)$ , com  $\|\cdot\|$  representando uma norma [GHOSH & NAG, 2001].

A norma euclidiana (ou norma  $L_2$ ), que pode ser calculada por (33), é tipicamente utilizada em interpolações, pois minimiza os erros [POWELL, 1985].

$$\|r - \mu_c\| = \sqrt{(r - \mu_c)^2} \quad (33)$$

Contudo, as FBRs são, na verdade, uma família de funções, dentre as quais algumas escolhas populares para  $\phi$  são exemplificadas na Figura 16.

Percebe-se que as quatro funções do exemplo são simétricas ao redor do centro (na Figura 16, centro com valor 50). Outra característica a se observar é que são monotonicamente crescentes ou decrescentes ao se afastarem do ponto central e no caso especial da função gaussiana é necessário especificar o desvio padrão (na Figura 16.c, valor 45), que ajusta a suavidade da função.



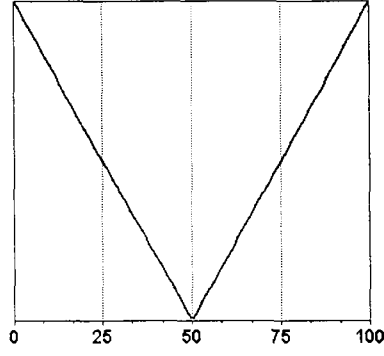
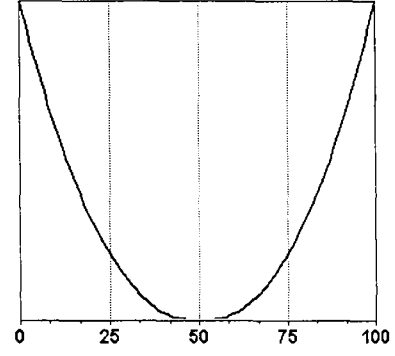
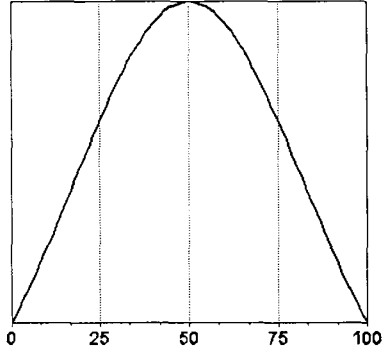
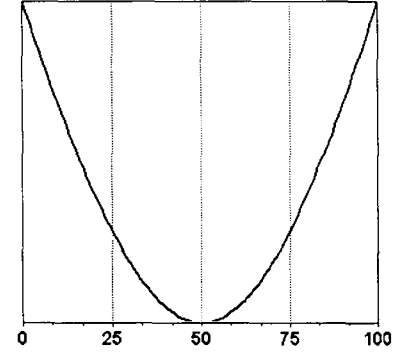
(a)  $\phi(r) = r$  (linear)(b)  $\phi(r) = r^2 \log(r)$  (thin-plate spline)(c)  $\phi(r) = e^{-r^2 / \sigma^2}$  (gaussiana)(d)  $\phi(r) = \sqrt{r^2 + \mu_c^2}$  (multiquadrática)

Figura 16 - Exemplos de funções radiais

As aproximações usando FBRs foram introduzidas no trabalho de [POWELL, 1985]. São como elementos de aproximação que as FBRs são utilizadas neste trabalho, isto é, na etapa de treinamento, os limiares determinados por sujeitos humanos são aproximados em função dos atributos de algumas imagens; e na etapa de teste, as respostas são generalizadas a partir dos atributos usados no treinamento.

As aproximações com FBRs podem ser vistas como um mapeamento de dados de entrada com dimensionalidade  $d$  em um espaço de saída unidimensional, ou seja,  $R^d \rightarrow R$ . Desse modo, o problema passa a ser: encontrar uma função  $F$  que represente o mapeamento de um conjunto de pontos de entrada  $S = \{\bar{s}_1, \bar{s}_2, \dots, \bar{s}_n\}$  em um espaço  $R^d$  de  $d$  dimensões, para um conjunto de pontos de saída  $T = \{t_1, t_2, \dots, t_n\}$  em um espaço  $R$ . Essa transformação deve satisfazer a condição de interpolação (34).

$$F(\bar{s}_i) = t_i, \quad i = 1, 2, \dots, n \quad (34)$$

Segundo [GHOSH & NAG, 2001], a solução para este problema pode ser a interpolação com uma base de  $n$  FBRs, utilizando o sistema linear em (35).

$$F(\bar{s}) = \sum_{i=1}^n w_i \phi(\|\bar{s} - \bar{s}_i\|) \quad (35)$$

O sistema linear em (35) pode ser visto de forma matricial como (36).

$$\begin{bmatrix} \bar{\mathbf{f}}_s \\ f(\bar{s}_1) \\ f(\bar{s}_2) \\ \vdots \\ f(\bar{s}_n) \end{bmatrix} = \begin{bmatrix} 0 & \phi(\|\bar{s}_1 - \bar{s}_2\|) & \cdots & \phi(\|\bar{s}_1 - \bar{s}_n\|) \\ \phi(\|\bar{s}_2 - \bar{s}_1\|) & 0 & \cdots & \phi(\|\bar{s}_2 - \bar{s}_n\|) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(\|\bar{s}_n - \bar{s}_1\|) & \phi(\|\bar{s}_n - \bar{s}_2\|) & \cdots & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \quad (36)$$

No sistema linear (36), o vetor  $\bar{\mathbf{f}}_s = [f(\bar{s}_1), f(\bar{s}_2), \dots, f(\bar{s}_n)]^T$  representa as respostas desejadas. A matriz  $\Phi$  é o conjunto de FBRs aplicadas às normas dos vetores de diferença entre todas as combinações de duas amostras  $\bar{s}$ . O vetor  $\bar{\mathbf{w}} = [w_1, w_2, \dots, w_n]^T$  é o vetor de pesos, ou seja, as incógnitas do sistema com  $n$  amostras. Cada um dos vetores de entrada  $\bar{s}$  tem dimensão  $d_s$ , ou seja, a quantidade de características usadas para representar cada amostra, resultando como aproximação uma curva ( $d_s = 1$ ), ou uma superfície ( $d_s = 2$ ) ou ainda uma hiper-superfície ( $d_s > 2$ ).

A matriz  $\Phi$  é quadrada e de ordem  $n$ , já que todos os elementos do conjunto de entradas  $S = \{\bar{s}_i | i = 1, 2, \dots, n\}$  são utilizados. Cada uma das amostras funcionará como centro de uma FBR.

Para efeito de simplicidade, permite-se escrever a equação (36) de maneira reduzida, como em (37).

$$F_{\bar{s}} = \Phi \bar{\mathbf{w}} \quad (37)$$

Durante o treinamento, a informação buscada é o vetor de pesos  $\bar{\mathbf{w}}$ . Desse modo, assume-se que a matriz  $\Phi$  seja não-singular e que então exista a sua inversa  $\Phi^{-1}$ . Com esta suposição o vetor de pesos  $\bar{\mathbf{w}}$  pode ser obtido pela solução do sistema linear, como mostrado em (38).

$$\vec{w} = \Phi^{-1} F_{\vec{s}} \quad (38)$$

Torna-se evidente garantir que a matriz  $\Phi$  seja inversível. Esta condição pode ser alcançada pelo teorema de Micchelli, o qual é aplicado para uma ampla classe de FBRs, incluindo gaussianas e thin-plate splines [HAYKIN, 2001].

A única condição que deve ser observada para que uma matriz de FBRs seja não singular é que os pontos do conjunto de entrada  $S = \{\vec{s}_i | i = 1, 2, \dots, n\}$  não sejam muito próximos uns dos outros, independente do tamanho de amostras de entrada  $n$  ou da dimensão de cada amostra  $d_s$ .

Uma vez solucionado o sistema, os valores dos pesos estão no vetor  $\vec{w}$ . Este vetor pode agora ser utilizado para encontrar as respostas desejadas de um outro conjunto de amostras (teste), de acordo com as características apresentadas por cada uma delas. Este processo representa a generalização das respostas de novas amostras ( $\vec{x}$ ) em função de suas características, conforme (39).

$$F_{\vec{x}} = \Phi(\|\vec{x} - \vec{s}\|) \vec{w} \quad (39)$$

A equação (39) pode ser escrita na forma matricial como segue:

$$\begin{bmatrix} f(\vec{x}_1) \\ f(\vec{x}_2) \\ \vdots \\ f(\vec{x}_m) \end{bmatrix} = \begin{bmatrix} \phi(\|\vec{x}_1 - \vec{s}_1\|) & \phi(\|\vec{x}_1 - \vec{s}_2\|) & \cdots & \phi(\|\vec{x}_1 - \vec{s}_n\|) \\ \phi(\|\vec{x}_2 - \vec{s}_1\|) & \phi(\|\vec{x}_2 - \vec{s}_2\|) & \cdots & \phi(\|\vec{x}_2 - \vec{s}_n\|) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(\|\vec{x}_m - \vec{s}_1\|) & \phi(\|\vec{x}_m - \vec{s}_2\|) & \cdots & \phi(\|\vec{x}_m - \vec{s}_n\|) \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \quad (40)$$

Como pode ser observado, a nova matriz  $\Phi$  contém  $m \times n$  elementos, sendo  $n$  o número de amostras utilizadas durante o treinamento e  $m$  o número de amostras utilizadas na generalização (teste). Novamente são usados todos os elementos do conjunto  $S = \{\vec{s}_i | i = 1, 2, \dots, n\}$ , agora centros das FBRs.

Na equação (40), o vetor  $\vec{f}_{\vec{x}} = [f(\vec{x}_1), f(\vec{x}_2), \dots, f(\vec{x}_m)]^T$  representa o conjunto de respostas desejadas para cada uma das  $m$  amostras. Cada elemento, por sua vez, é descrito por um vetor de características  $\vec{x}_j$  com dimensão igual ao vetor de amostras

de treinamento  $d_s$ . A matriz  $\Phi$ , agora, representa o conjunto de FBRs aplicadas às normas das diferenças entre os vetores de características  $\bar{s}_i$  e  $\bar{x}_j$ , das amostras de treinamento e teste, respectivamente. O vetor  $\bar{w} = [w_1, w_2, \dots, w_n]^T$  é o mesmo vetor de pesos obtido como solução do sistema linear em (38). As respostas generalizadas são determinadas aplicando então (39).

Segundo Bishop [BISHOP, 1995], nem sempre a interpolação exata dos dados é desejada, uma vez que os dados de entrada podem estar contaminados por ruído ou por idiossincrasias e, dessa forma, podem prejudicar a generalização dos dados. Uma boa alternativa é desenvolver a generalização pela média entre o ruído e os dados e assim gerar uma superfície mais suave que generalize melhor a informação.

Uma outra limitação da interpolação exata é a necessidade de uma base com um número de funções igual ao número de amostras utilizadas, o que, dependendo do tamanho do conjunto de entrada, torna este processo computacionalmente muito caro, podendo até inviabilizar o seu uso.

Uma forma de conseguir a generalização evitando estas restrições é usar uma rede de FBRs (RFBR), que melhora a generalização suavizando a função de interpolação. As RFBRs oferecem também a vantagem de exigir um número de funções muito menor que o número de amostras de entrada. A escolha das amostras usadas como centros das RFBRs faz parte do treinamento, caracterizada por Haykin [HAYKIN, 2001] como estratégia de aprendizagem.

Uma rede de funções de base radial genérica é mostrada na Figura 17. Nota-se que cada uma das  $n$  amostras do conjunto de entradas  $S = \{\bar{s}_i | i = 1, 2, \dots, n\}$  alimenta cada uma das  $k$  funções da base, as quais são linearmente combinadas com os pesos  $w_k$ , resultando na saída da rede  $f(\bar{s})$ .

O bias elimina a necessidade da interpolação passar pela origem do espaço  $d_s$  dimensional [BISHOP, 1995]. Por exemplo, se considerarmos a interpolação em um espaço bidimensional, temos uma linha e então o bias define a distância perpendicular entre a linha e sua origem no plano cartesiano  $(0, 0)$ .

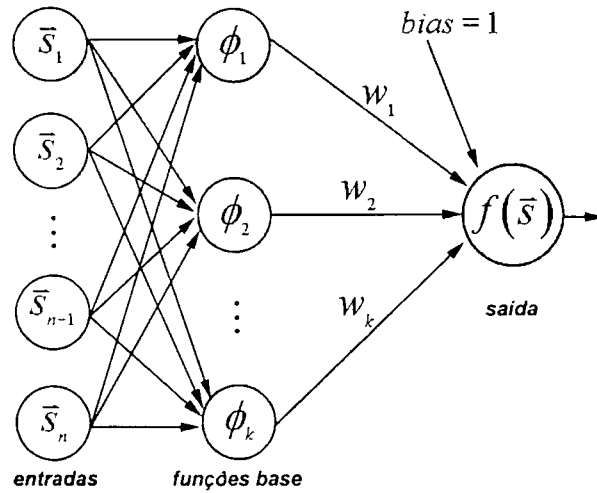


Figura 17 - Rede de funções de base radial tradicional [HAYKIN, 2001].

Usando esta nova arquitetura e considerando a inclusão do bias, as respostas são obtidas pela equação (41).

$$F(\bar{s}) = \sum_{i=1}^k w_i \phi_i(\|\bar{s} - \bar{\mu}_i\|) + b \quad (41)$$

A qual pode ser escrita na forma matricial como em (42).

$$\begin{bmatrix} f(\bar{s}_1) \\ f(\bar{s}_2) \\ \vdots \\ f(\bar{s}_n) \end{bmatrix} = \begin{bmatrix} \phi_1(\|\bar{s}_1 - \bar{\mu}_1\|) & \phi_2(\|\bar{s}_1 - \bar{\mu}_2\|) & \cdots & \phi_k(\|\bar{s}_1 - \bar{\mu}_k\|) & 1 \\ \phi_1(\|\bar{s}_2 - \bar{\mu}_1\|) & \phi_2(\|\bar{s}_2 - \bar{\mu}_2\|) & \cdots & \phi_k(\|\bar{s}_2 - \bar{\mu}_k\|) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi_1(\|\bar{s}_n - \bar{\mu}_1\|) & \phi_2(\|\bar{s}_n - \bar{\mu}_2\|) & \cdots & \phi_k(\|\bar{s}_n - \bar{\mu}_k\|) & 1 \end{bmatrix}_{n \times k+1} \times \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \\ b \end{bmatrix} \quad (42)$$

A estratégia de aprendizagem usada neste trabalho é classificada como uma seleção auto-organizada de centros, usando para isto o algoritmo de agrupamento  $k$ -médias [CONSULARO, 2000], descrito na seção 2.3.5. O  $k$ -médias determinará o conjunto de centros  $M = \{\bar{\mu}_i | i=1,2,\dots,k\}$  pelo agrupamento das amostras do conjunto de entradas  $S = \{\bar{s}_i | i=1,2,\dots,n\}$  em  $k$  grupos, sendo cada um deles tão homogêneos quanto possíveis. Para cada um dos grupos é obtida a média e esta será o centro escolhido. O objetivo é selecionar os centros mais representativos.

Como se pode perceber, a matriz de funções  $\Phi$  deixa de ser quadrada, pelo fato de possuir mais dados de entrada do que FBRs, o que torna o problema

superdeterminado [HAYKIN, 2001] e conseqüentemente não existe uma inversa única para a matriz. Desse modo, é usada a solução da pseudo-inversa [BROOMHEAD & LOWE, 1988] na determinação das respostas desejadas. Sendo que  $\Phi^+$  representa a pseudo-inversa da matriz  $\Phi$ , como mostra a equação (43).

$$\Phi^+ = (\Phi^T \Phi)^{-1} \Phi^T \quad (43)$$

Conseqüentemente, a aproximação que antes era obtida por (37) passa a ser obtida por (44).

$$F_{\bar{s}} = \Phi \bar{w} \quad (44)$$

E assim, os pesos antes obtidos pela equação (38) passam a ser determinados pela equação (45).

$$\bar{w} = \Phi^+ F_{\bar{s}} \quad (45)$$

Uma vez encontrados os pesos, a generalização para um outro conjunto de amostras (teste) é similar. Convém observar que o conjunto de centros selecionados durante o treinamento  $M = \{\bar{\mu}_i | i = 1, 2, \dots, k\}$ , continua sendo usado na generalização como mostra a equação (46).

$$\begin{bmatrix} f(\bar{x}_1) \\ f(\bar{x}_2) \\ \vdots \\ f(\bar{x}_m) \end{bmatrix} = \begin{bmatrix} \phi_1(\|\bar{x}_1 - \bar{\mu}_1\|) & \phi_2(\|\bar{x}_1 - \bar{\mu}_2\|) & \cdots & \phi_k(\|\bar{x}_1 - \bar{\mu}_k\|) & 1 \\ \phi_1(\|\bar{x}_2 - \bar{\mu}_1\|) & \phi_2(\|\bar{x}_2 - \bar{\mu}_2\|) & \cdots & \phi_k(\|\bar{x}_2 - \bar{\mu}_k\|) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi_1(\|\bar{x}_m - \bar{\mu}_1\|) & \phi_2(\|\bar{x}_m - \bar{\mu}_2\|) & \cdots & \phi_k(\|\bar{x}_m - \bar{\mu}_k\|) & 1 \end{bmatrix}_{m \times k+1} \times \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \\ b \end{bmatrix} \quad (46)$$

Por fim, usando os pesos e centros encontrados durante o treinamento as respostas desejadas são obtidas generalizando os dados de teste em função de suas características de entrada. Esta generalização é obtida aplicando-se (47).

$$F(\bar{x}) = \sum_{i=1}^k w_i \phi(\|\bar{x} - \bar{\mu}_i\|) + b \quad (47)$$

Segundo Orr [ORR, 1996], uma característica importante e útil das RFBRs, é permitir incluir e retirar funções da base, bem como incluir ou excluir características nas amostras de entrada  $d_s$  sem a necessidade de treinar novamente toda a rede. Esse aspecto é especialmente útil na seleção de características.

## 2.5 Psicofísica

A psicofísica, segundo Gordon [GORDON, 1995], não é uma teoria da percepção, mas sim um conjunto de técnicas, com o objetivo de medir a sensibilidade e os limiares da percepção.

A psicofísica teve origem com os trabalhos de Weber e Fechner. Em 1834, o fisiologista E. H. Weber aplicando experimentos com levantamento de pesos, percebeu a mínima diferença perceptível e o modo como ela se alterava com o aumento e redução dos pesos.

Weber, tentando relacionar o estímulo e a percepção, descobriu que se uma massa inicial com, por exemplo, 100 gramas, for pouco a pouco incrementado, haverá a percepção desse incremento apenas quando for atingida uma determinada diferença. Se esta diferença for de, por exemplo, 20 gramas, esta diferença é chamada JND (do inglês, “*Just Noticeable Difference*”), ou seja, a diferença perceptível [GORDON, 1995].

Dessa forma, o que Weber concluiu é que a JND não é obtida pela adição de um valor fixo ao estímulo inicial, mas sim pela razão entre o estímulo inicial e a JND. Este resultado ficou conhecido como lei de Weber (48).

$$\frac{\Delta S}{S} = K \quad (48)$$

Na equação (48) o  $\Delta S$  representa a variação do estímulo que foi perceptível a partir do estímulo inicial  $S$ , e  $K$  é uma constante para uma modalidade específica de percepção (frio, sabor, odor, visual, etc.). A relação de várias constantes pode ser encontrada em [STEVENS, 1986].

Gustav T. Fechner, em 1877 observando o trabalho desenvolvido por Weber, desejava criar unidades de percepção, as quais poderiam realmente medir a sensação e não apenas compará-las umas com as outras. Fechner incrementou o trabalho de Weber, adicionando uma nova característica. Ele propôs que para cada unidade de tempo, uma JND é adicionada ao estímulo e com isso a sensação (percepção) aumenta num salto de tamanho constante [STEVENS, 1986].

A combinação dos trabalhos desenvolvidos por Weber e Fechner sugere uma lei logarítmica para o crescimento das sensações como uma função do estímulo, e que proporções iguais entre estímulos produzem sensações iguais. A equação (49) representa a função de estímulo, cujo comportamento pode ser observado na Figura 18, com  $S$  representando a sensação; a constante  $K$  representando a mesma constante da lei de Weber e  $E$  representando o estímulo.

$$S = K \log(E) \quad (49)$$

Esta equação caracteriza a lei Weber-Fechner. A união das pesquisas desenvolvidas por Weber e Fechner ficou caracterizada como sendo o princípio da lei psicofísica, que define a sensação como uma função logarítmica do estímulo.

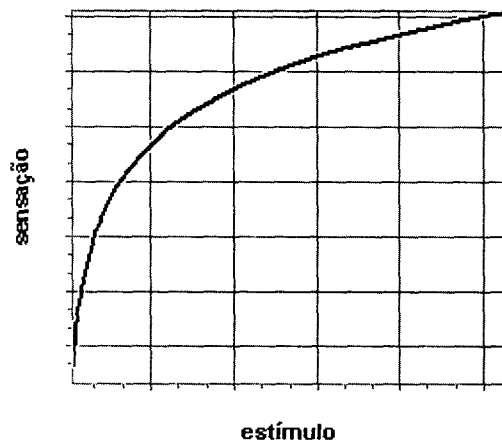


Figura 18 - Lei de Weber e Fechner

Um experimento psicofísico se caracteriza pela presença dos seguintes fatores:

- a) O experimentador varia (manipula) algum fator, isto é, o estímulo apresentado  $E$ .
- b) O experimentador mantém as outras condições constantes.
- c) O experimentador verifica o efeito da variação sobre o fenômeno que está observando, ou seja, sobre a sensação  $S$ .



No experimento psicofísico existem variáveis dependentes e independentes. Uma variável independente é uma condição estabelecida pelo experimentador (estímulo apresentado). A variável dependente é o comportamento do sujeito que está sendo exposto ao estímulo. Recebe este nome porque seu valor depende diretamente do estímulo apresentado.

Existem três métodos clássicos utilizados pela psicofísica para encontrar o limiar da percepção: o método dos limites, o método constante e o método de ajuste [GORDON, 1995].

O método dos limites ou mudanças mínimas consiste na seleção de um estímulo padrão, por exemplo, o brilho de uma lâmpada. Esse brilho é sistematicamente avaliado pela comparação com outras lâmpadas, inicialmente mais brilhantes. A intensidade é reduzida até que o observador perceba que ela está menos brilhante que o estímulo padrão, quando então, o teste é finalizado.

Durante o teste são anotadas as intensidades limites que o observador conseguiu identificar quando o estímulo estava mais brilhante, com brilho igual e menos brilhante que o estímulo padrão.

No método constante são apresentados: um estímulo padrão e outros diversos estímulos. O observador deve dizer se este último é maior ou menor que o padrão. A proporção de respostas corretas é armazenada, a qual determinará o resultado do método.

O mais utilizado dentre eles é o método de ajuste. Nele é apresentado um estímulo padrão e então se pede ao observador para alterar outro estímulo de modo que fique igual ou muito próximo ao estímulo padrão. Por exemplo, uma imagem limiarizada num limiar  $T$  pode servir como estímulo padrão. A seguir, é solicitado que o observador limiarize outra imagem com o objetivo de encontrar o valor  $T$ .

### 3 METODOLOGIA

O desenvolvimento de um modelo baseado na percepção visual envolve algumas etapas: 1) a definição de características; 2) a elaboração de um experimento psicofísico; 3) a escolha de um conjunto de treinamento e testes; 4) a elaboração do modelo; e 5) a validação da aplicação do modelo. É comum encontrar métodos de validação que exigem um conjunto único para treinamento e testes. Neste caso as etapas 3 e 4 se fundiriam em uma única etapa.

Neste capítulo será descrito como cada uma destas etapas foi implementada, mostrando as escolhas assumidas e suas consequências.

#### 3.1. A Definição de Características

Como o objetivo deste trabalho é investigar os relacionamentos entre a percepção visual e os resultados obtidos por métodos de limiarização automática, é necessário definir características utilizadas por estes métodos de limiarização. Como a própria revisão cita, a entropia e a variância estão presentes em grande parte deles [OTSU, 1978], [PUN, 1980], [JOHANNSEN & BILLE, 1982], [KAPUR et al., 1985], [ABUTALEB, 1989], [BEGHDADI et. al., 1995], [BRINK, 1996], [BRINK & PENDOCK, 1996], [FLEURY et. al., 1996].

#### 3.2. A Elaboração do Experimento Psicofísico

O experimento psicofísico<sup>6</sup> deve coletar informações fornecidas por sujeitos humanos instruídos a escolher um limiar para diversas imagens. Estas imagens foram escolhidas de forma a tentar cobrir uma escala de entropia e variância. Os dados obtidos dessa coleta são armazenados em um banco de dados para que possam ser posteriormente analisados e modelados.

---

<sup>6</sup> <http://www.mycgiserver.com/~fabricio/limiarimg.html>

Os estímulos (variável independente) são imagens em tons de cinza e a variável dependente (determinada pelo sujeito), um limiar escolhido para a imagem apresentada.

Na execução do experimento, é solicitado ao sujeito sua identificação: nome e e-mail. Uma barra deslizante é disponibilizada para que se possa determinar um limiar de nível de intensidade na imagem. Cada sujeito foi instruído a encontrar entre as intensidades de cinza um limiar de acordo com a melhor separação entre frente e fundo. É permitido ao sujeito percorrer entre as intensidades de cinza e visualizar em tempo real a imagem limiarizada pela intensidade selecionada.

Após o sujeito escolher o limiar que o agrada, é permitido a ele remover regiões de pixels que avalie não corresponder ao objeto, usando uma ferramenta borracha. Tendo feito isso para uma imagem, é solicitado que grave sua escolha num banco de dados. Junto às informações do usuário já estão armazenados os atributos (seção 2.1) de cada imagem e os limiares atribuídos pelos métodos automáticos: máxima entropia [KAPUR et al., 1985], máxima variância [OTSU, 1978] e “maxmin” entropia [BRINK, 1996]. Estas informações servirão como entrada para um eventual modelo.

A Figura 19 mostra a janela inicial do aplicativo do experimento, que exibe algumas instruções ao sujeito. Estas instruções pedem para que o sujeito clique uma imagem em tons de cinza e no local correspondente a imagem limiarizada é exibida uma imagem com pixels distribuídos aleatoriamente (ruído branco), para que o sujeito não sofra a influência de um estado inicial da barra deslizante, que neste caso ficará inicialmente oculta.



Figura 19- Tela inicial do experimento. Nesta tela são mostradas as instruções para que o sujeito clique inicialmente na imagem (da esquerda) antes de escolher algum limiar.

Um total de 137 alunos de graduação, do segundo e terceiro período, do curso de Tecnologia em Informática do Centro Federal de Educação Tecnológica do Paraná (CEFET-PR - unidade de Cornélio Procopio) foram submetidos ao experimento espontaneamente, sem nenhuma recompensa. A cada sujeito foram apresentadas 12 imagens diferentes, escolhidas aleatoriamente, de um conjunto de 110 imagens (estímulos). O conteúdo das 110 imagens foi diverso, abrangendo faces, paisagens, carros e casas. Os tamanhos das imagens foram também distintos, como uma forma de não influenciar nos resultados.

Cada sujeito demorou, em média, 37 minutos para terminar o experimento com as 12 imagens. A aplicação do experimento ocorreu nos laboratórios de informática da instituição, os quais são iluminados por lâmpadas fluorescentes de 40 watts do tipo luz do dia, tendo uma iluminância de 210 lux no ponto de trabalho de cada aluno. O lux é uma medida do fluxo luminoso por metro quadrado. A tabela padrão de iluminâncias mínimas [FILHO, 2001] divide as iluminâncias por atividade e categoria. Na atividade escolas e categoria laboratórios o valor padrão é de 200 lux, dessa forma, a iluminância no ponto de trabalho foi adequada para a realização dos experimentos.

Nos laboratórios para cada computador existe uma mesa e uma cadeira, assim durante o experimento cada sujeito permaneceu sentado à frente do monitor, a uma distância média de 60 centímetros entre o seu olho e o dispositivo de vídeo, que fica a uma altura de 1 metro do chão. Os experimentos ocorreram entre 12 de novembro de 2002 e 18 de dezembro de 2002.

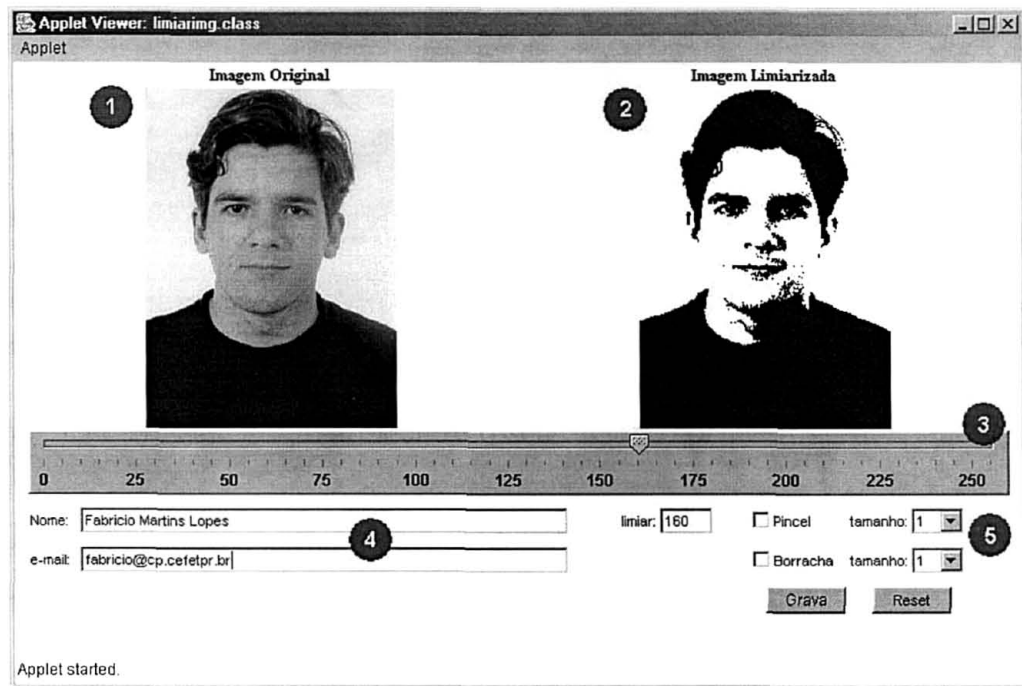


Figura 20- Captura da janela com os componentes do aplicativo que implementa o experimento. A imagem original (1), a imagem limiarizada (2) pelo limiar escolhido pelo sujeito usando a barra deslizante (3). As caixas de entrada para identificação (4), as configurações da borracha para remover pixels que o sujeito avalie não pertencer ao objeto e dois botões: um para gravar o resultado e outro para reiniciar o experimento para aquela imagem (5).

As informações coletadas foram armazenadas em um banco de dados, sobre uma arquitetura cliente-servidor. A Figura 20 mostra a parte visível ao usuário (applet), que representa o cliente. As informações atribuídas pelo usuário são transmitidas de forma serial para o servidor (servlet) em uma operação transacional. O servlet espera até receber toda a informação para armazená-la em um banco de dados.

As motivações no uso de um ambiente web para a coleta dos dados foram: (1) a flexibilidade nas alterações das versões do experimento; (2) a potencialidade em atingir grandes audiências em locais diversos; e (3) a capacidade de ocorrência simultânea de experimentos. Nem todo experimento pode ser implementado sem um

controle ambiental, como é o caso dos experimentos baseados nos métodos dos limites e também no método constante [GORDON, 1995].

Estas motivações levaram à decisão de desenvolver o ambiente utilizando a linguagem Java, utilizando as tecnologias: applets e servlets, que possibilitam a implementação da arquitetura cliente-servidor em ambiente web.

Java apresenta como uma de suas principais vantagens a portabilidade para vários sistemas operacionais e sua facilidade no tratamento de aplicações destinadas a web (applets e servlets). Estas vantagens são, em parte, decorrentes de seus compromissos com a análise e programação orientação a objetos, com a segurança de seu código e da máquina virtual sobre a qual todas as aplicações Java executam.

A compilação de um código Java (arquivo.java), gera um código chamado byte-code (arquivo.class), isto é, um código intermediário entre a linguagem de máquina e a linguagem de alto nível. O byte-code é interpretado por uma máquina virtual JVM (Java Virtual Machine), que é disponibilizada para uma grande diversidade de sistemas operacionais. Deste modo, os aplicativos desenvolvidos podem ser executados (interpretados) em várias plataformas diferentes.

Um outro fator de fundamental importância para a escolha da linguagem Java neste trabalho é a implementação da biblioteca JAI (Java Advanced Image). Esta biblioteca contém métodos típicos de processamento de imagens, de representações de imagens e também de leitura e escrita de imagens em vários formatos.

Como ambiente de desenvolvimento, foi utilizado o Forté para Java, versão 3, edição para a comunidade, hoje distribuído com o nome StudioOne. A versão do compilador foi a 1.4.0. Todos estes softwares (ambiente de desenvolvimento e compilador) são de uso livre, bem como a biblioteca JAI, a qual foi usada a versão 1.1.1\_01.

### **3.3. A Elaboração do Modelo**

O desenvolvimento de um modelo que possa, baseado em características globais de um conjunto de imagens e nas respostas de sujeitos a um experimento com

estas mesmas imagens, responder de acordo com as características recebidas requer um método de generalização.

Existem diversos métodos de generalização. Sejam interpoladores ou regressões [MOORE, 2000] agrupadores ou classificadores [DUDA et. al., 1998], [THEODORIDIS & KONSTANTINOS, 1999], todos exigem um conjunto de dados como referência para uma próxima entrada. Contudo, agrupadores e classificadores não geram explicitamente um modelo, mas sim uma resposta de grupo ou classe. Neste tipo de generalização, a ordem em que os dados são apresentados pode alterar profundamente as respostas.

Como o desenvolvimento de um modelo de percepção visual faz parte do conjunto de objetivos deste trabalho, é conveniente que o método que resolva o modelo seja robusto, ou seja, um método cuja natureza dos dados não inviabiliza numérica ou algebricamente os resultados. Além disso, o método deve permitir processos de validação com os quais se identifique as melhores respostas. Outra característica importante é a flexibilidade, isto é, o método não pode impor restrições rígidas, tal como uma regressão linear, de modo a induzir um suposto conhecimento ao modelo.

### **3.3.1. A Interpolação Usando Thin-Plate Splines**

Inicialmente foi implementado um método de generalização baseado nas *Thin-Plate Splines* (TPS). Neste desenvolvimento usou-se as respostas colhidas de sujeitos humanos, conforme o experimento descrito na seção 3.2, e uma medida de entropia de cada uma das imagens. O modelo é então representado pela curva interpolada por um conjunto de pontos. Cada ponto corresponde a uma imagem com uma entropia e um limiar médio escolhido pelos sujeitos. O limiar de uma nova imagem pode ser obtido aplicando o valor de sua entropia ao modelo.

É claro que existe uma infinidade de curvas passando pelos mesmos pontos e, portanto, é necessário impor uma restrição para que se determine o modelo. Os modelos baseados em TPS são soluções que minimizam a energia interna da curva.

Esta restrição impõe uma suavidade à curva (e eventualmente às superfícies) impedindo, ou pelo menos reduzindo, a ocorrência de descontinuidades ou sobressaltos [SANDWELL, 1987]. A TPS é uma família de funções de base radial (FBR) da forma:  $f(r) = r^n$  ou  $f(r) = r^n \log(r)$ , para  $n > 0,5$ . Quando  $n = 2$  a suavidade da TPS é máxima [POWELL, 1985].

Utilizando TPS, foi desenvolvida uma classe Java tendo entre seus métodos um que é capaz de gerar uma curva descrevendo parametricamente o limiar humano como uma função da entropia. Desta maneira, os parâmetros da função definem um modelo entrópico da limiarização realizada por humanos. Do banco de dados é resgatado um conjunto de imagens de teste, diferentes daquelas utilizadas no treinamento, e suas respectivas entropias previamente calculadas e armazenadas. Aplicada a interpolação, o limiar para a imagem de teste é obtido em função de sua entropia, ou seja, um mapeamento  $R^1 \rightarrow R^1$ .

Na fase de treinamento cada imagem é representada por um vetor de características  $\bar{s}_k$ , que neste caso é unidimensional e cujo valor é a própria entropia da imagem. O valor  $k$  indexa cada uma das imagens. A função  $f(\bar{s})$  corresponde ao limiar escolhido por um sujeito humano, para cada imagem  $k$ . A função de base radial  $\phi$  é aplicada sobre todas as distâncias (ou diferenças) entre as entropias das imagens. Assim, o treinamento torna-se simplesmente resolver o sistema linear abaixo. A norma  $\|\cdot\|$  utilizada foi a euclidiana.

$$f(\bar{s}) = \sum_{k=1}^n w_k \phi(\|\bar{s} - \bar{s}_k\|) + \bar{a} \times \bar{s} + b \quad (50)$$

Para este sistema linear é necessário ainda impor algumas restrições (51) para que a função modelada não se fixe à origem e nem aos limites, permitindo assim uma interpolação mais flexível e com menores erros.

$$\sum_{k=1}^n w_k \cdot s_k = 0 \quad \text{e} \quad \sum_{k=1}^n w_k = 0 \quad (51)$$

Por fim, o que se pode chamar de modelo é o conjunto de coeficientes  $\{w_k, \bar{a}, b\}$ , onde o vetor  $\bar{a}$  e o escalar  $b$  são, segundo Mees [MEES et. al., 1992], as imposições de um comportamento afim da interpolação.



A solução na forma matricial e genérica para  $d$  dimensões, ou seja, a interpolação afim com base radial em um espaço  $R^d$  pode ser melhor visualizada em (52).

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ a_1 \\ a_2 \\ \vdots \\ a_d \\ b \end{bmatrix} = \begin{bmatrix} 0 & \phi(\|\bar{s}_1 - \bar{s}_2\|) & \cdots & \phi(\|\bar{s}_1 - \bar{s}_n\|) & s_{11} & s_{12} & \cdots & s_{1d} & 1 \\ \phi(\|\bar{s}_2 - \bar{s}_1\|) & 0 & \cdots & \phi(\|\bar{s}_2 - \bar{s}_n\|) & s_{21} & s_{22} & \cdots & s_{2d} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi(\|\bar{s}_n - \bar{s}_1\|) & \phi(\|\bar{s}_n - \bar{s}_2\|) & \cdots & 0 & s_{n1} & s_{n2} & \cdots & s_{nd} & 1 \\ s_{11} & s_{21} & \cdots & s_{n1} & 0 & 0 & \cdots & 0 & 0 \\ s_{12} & s_{22} & \cdots & s_{n2} & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ s_{1d} & s_{2d} & \cdots & s_{nd} & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}^{-1} \times \begin{bmatrix} f(s_1) \\ f(s_2) \\ \vdots \\ f(s_n) \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad (52)$$

Para o caso  $d = 1$ ,  $\phi(r) = r^2 \log(r)$ , tendo como característica de imagens a entropia, e com o conjunto de 11 imagens mostrado na Tabela 1, com seus respectivos limiares coletados no experimento e respectivas entropias, o resultado da interpolação é ilustrado na Figura 21.

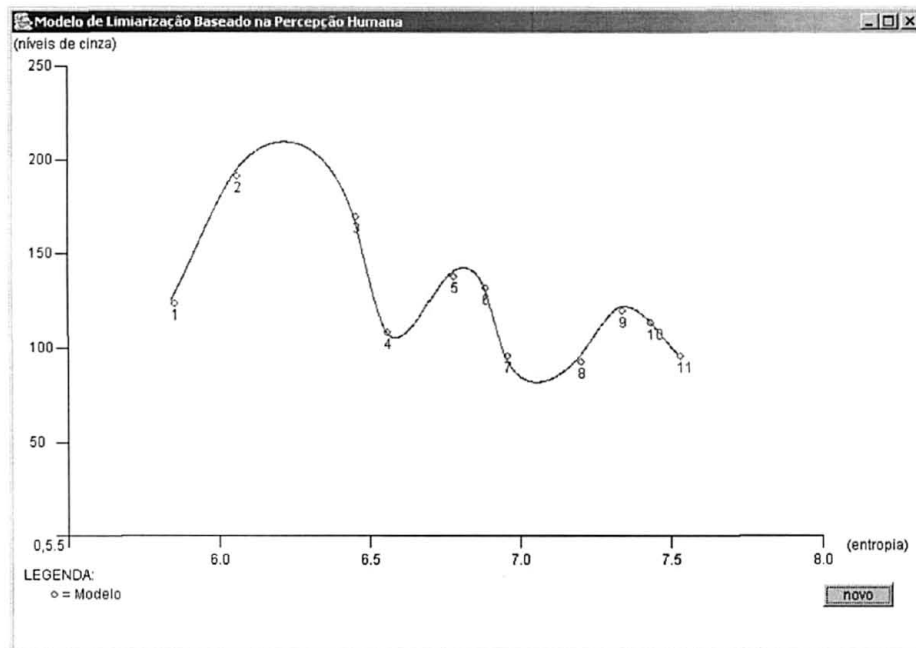





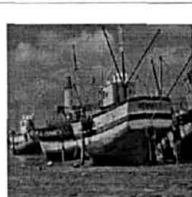


Figura 21 - Exemplo de Modelo de Limiarização usando a interpolação por base radial com TPS, Cada centro (imagem) está representado por um círculo numerado.

Inicialmente foram realizados alguns treinamentos com amostras variadas entre 6 e 18 imagens. Para cada treinamento foi usado um conjunto de teste, quase sempre de mesmo tamanho, tendo variação máxima de 3 imagens a mais ou a menos.

Os limiares foram então generalizados assumindo a curva interpolada como uma função da entropia.

identificação	Nome da Imagem	Imagem	Entropia	Limiar Humano
1	paisagem4		5,844	125
2	casa2		6,054	193
3	casa3		6,446	171
4	paisagem5		6,548	110
5	câmera		6,769	139
6	boat		6,878	133

7	paisagem7		6,946	97
8	zelda		7,191	94
9	lena		7,328	121
10	pimentões		7,422	115
11	coruja		7,519	97

Tabela 1 - Exemplo de imagens usadas durante a geração do modelo TPS e suas respectivas entropias e limiares humanos.<sup>7</sup>

Após seu desenvolvimento, o modelo foi aplicado, conforme mostra a equação (53), para a generalização de um conjunto de dados desconhecidos (teste), isto é, a aplicação do modelo deve retornar o limiar para cada imagem de acordo com a sua entropia. O teste tem o propósito de avaliar o desempenho do modelo gerado. Na

<sup>7</sup> Department of Computer Science - Princeton University [http://www.cs.princeton.edu/~dbradley/profile\\_images1.zip](http://www.cs.princeton.edu/~dbradley/profile_images1.zip). Signal and Machine Perception Laboratory – Department of Electrical Engineering – The Ohio State University. <http://sampl.eng.ohio-state.edu/~sampl/database.htm>. Signal and Image Processing Institute of the University of Southern California. <http://sipi.usc.edu/services/database/Database.html>

equação (53), o conjunto de teste com os valores de entropia para cada imagem é representado por  $x_j, j = 1, 2, \dots, m$ .

$$\begin{matrix} \vec{f}_x & & \Phi & & \vec{w} \\ \begin{bmatrix} f(\bar{x}_1) \\ f(\bar{x}_2) \\ \vdots \\ f(\bar{x}_m) \end{bmatrix} & = & \begin{bmatrix} \phi(\|\bar{x}_1 - \bar{s}_1\|) & \phi(\|\bar{x}_1 - \bar{s}_2\|) & \cdots & \phi(\|\bar{x}_1 - \bar{s}_n\|) \\ \phi(\|\bar{x}_2 - \bar{s}_1\|) & \phi(\|\bar{x}_2 - \bar{s}_2\|) & \cdots & \phi(\|\bar{x}_2 - \bar{s}_n\|) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(\|\bar{x}_m - \bar{s}_1\|) & \phi(\|\bar{x}_m - \bar{s}_2\|) & \cdots & \phi(\|\bar{x}_m - \bar{s}_n\|) \end{bmatrix} & \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} & 1 \\ x_{21} & x_{22} & \cdots & x_{2d} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{md} & 1 \end{bmatrix} & \times & \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ a_1 \\ a_2 \\ \vdots \\ a_d \\ b \end{bmatrix} \end{matrix} \quad (53)$$

O vetor  $\vec{w}$  é o modelo resultante da solução do sistema linear do treinamento. Cada elemento do vetor  $\vec{f}_x$  é uma função de  $x$  representando o limiar, isto é, a informação desejada. A matriz  $\Phi$  é a matriz de interpolação, cada elemento é uma função de base radial  $\phi$  e, como no treinamento,  $\|\cdot\|$  representa uma norma euclidiana.

Uma vez aplicada a generalização nos dados de teste, temos os limiares retornados pelo modelo em função das entropias de cada imagem. Um exemplo da generalização de um conjunto de teste contendo também 11 imagens, as quais são mostradas na Tabela 2, e suas entropias pode ser observado na Figura 22.

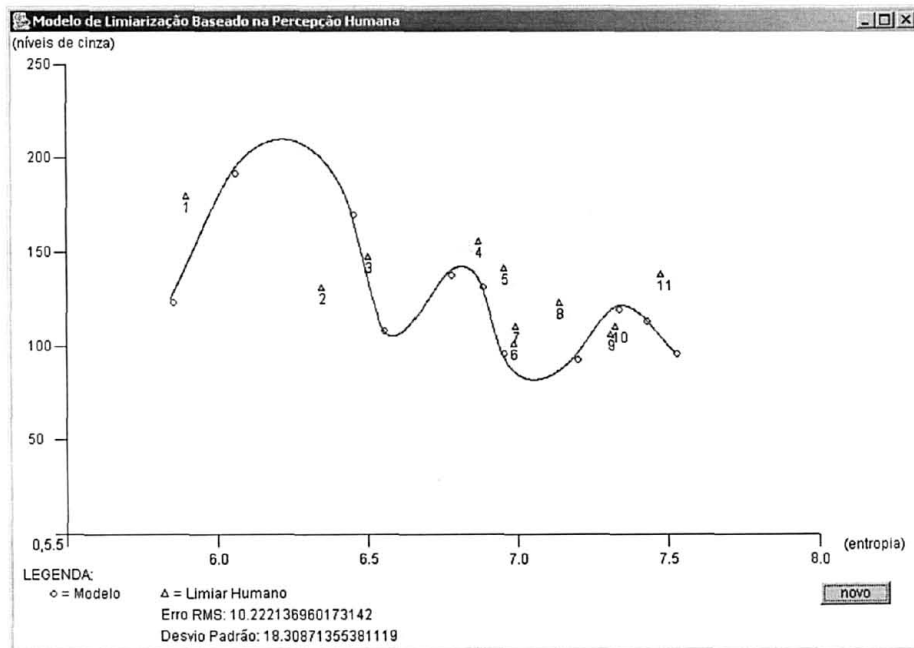


Figura 22- Aplicação do Modelo de Limiarização TPS. As respostas generalizadas pelo modelo estão representadas por triângulos numerados.

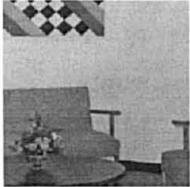



Uma vez aplicado o teste e encontrados os limiares para as imagens de teste, pode-se então avaliar o modelo calculando-se o erro médio quadrático (54).

$$Erro = \frac{\sqrt{(f(h_1) - f(x_1))^2 + \dots + (f(h_m) - f(x_m))^2}}{m} \quad (54)$$

Sendo:

$m$  representa o número de imagens de teste.

A raiz do erro médio quadrático (RMS) é calculado para as imagens do conjunto de teste, considerando a diferença entre o limiar retornado pelo modelo  $f(x)$  e o limiar atribuído previamente por um humano  $f(h)$ , isto é, a resposta correta. Um outro critério que é observado para analisar o modelo é o desvio padrão das diferenças entre  $f(x)$  e  $f(h)$ .

identificação	Nome da Imagem	Imagem	Entropia	Limiar Humano	Limiar TPS
1	quarto		5,887	181	143
2	paisagem3		6,336	133	199
3	paisagem1		6,493	149	137
4	paisagem6		6,861	157	135








5	paisagem2		6,945	143	93
6	6101		6,978	103	87
7	borboleta2		6,984	112	85
8	mandrill		7,130	125	86
9	goldhill		7,301	108	118
10	barb		7,317	112	120
11	tulipa2		7,468	140	105

Tabela 2 - Exemplo de imagens teste e suas respectivas entropias, limiares humanos e limiares retornados pelo modelo.<sup>8</sup>

<sup>8</sup> Department of Computer Science - Princeton University [http://www.cs.princeton.edu/~dbradley/profile\\_images1.zip](http://www.cs.princeton.edu/~dbradley/profile_images1.zip)

A interpolação TPS se mostrou eficiente na construção do modelo. No entanto, sempre que um novo dado de treinamento ou um novo atributo for incluído no modelo, um novo treinamento deve ocorrer. Cada treinamento envolve uma inversão de matrizes cujo tamanho é proporcional ao número de amostras de treinamento.

Estas limitações foram resolvidas elaborando-se um modelo mais flexível, usando ainda as funções de base radial TPS, mas agora em uma rede de funções de base radial (RFBR) TPS.

O conjunto de imagens utilizado no experimento psicofísico foi então utilizado para validar o modelo de RFBR, inicialmente com uma única característica: a entropia. Outras características de imagens foram utilizadas para compará-las com métodos de limiarização automática já conhecidos e também com o intuito de observar o comportamento dos modelos com estas características.

### 3.3.2. Generalização com RFBR

O modelo de interpolação afim com funções de base radial TPS inspirou o uso de uma rede de funções de base radial (RFBR), e assim, o treinamento passou a ser realizado desvinculando-se o número de centros de cada função do número de amostras do conjunto de treinamento. Deste modo na equação (55), pode-se observar que a função de base radial  $\phi$  não é mais calculada sobre as distâncias entre todas as amostras, mas agora entre as amostras e um centro  $\bar{\mu}_i$  da base radial da função.

$$F(\bar{s}) = \sum_{i=1}^k w_i \phi(\|\bar{s} - \bar{\mu}_i\|) + b \quad (55)$$

A equação (55) pode ser escrita na forma matricial como:

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \\ b \end{bmatrix} = \begin{bmatrix} \phi_1(\|\bar{s}_1 - \bar{\mu}_1\|) & \phi_1(\|\bar{s}_2 - \bar{\mu}_1\|) & \cdots & \phi_1(\|\bar{s}_n - \bar{\mu}_1\|) \\ \phi_2(\|\bar{s}_1 - \bar{\mu}_2\|) & \phi_2(\|\bar{s}_2 - \bar{\mu}_2\|) & \cdots & \phi_2(\|\bar{s}_n - \bar{\mu}_2\|) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_k(\|\bar{s}_1 - \bar{\mu}_k\|) & \phi_k(\|\bar{s}_2 - \bar{\mu}_k\|) & \cdots & \phi_k(\|\bar{s}_n - \bar{\mu}_k\|) \\ 1 & 1 & \cdots & 1 \end{bmatrix}^{+} \times \begin{bmatrix} f(\bar{s}_1) \\ f(\bar{s}_2) \\ \vdots \\ f(\bar{s}_n) \end{bmatrix} \quad (56)$$

$k+1 \times n$

A inclusão do termo  $b$  (bias) no vetor de pesos  $\bar{w}$  tem o mesmo sentido dos termos usados:  $b$  e  $\bar{a}$ , na interpolação exata. A dimensão do conjunto de amostras de entrada pode agora variar e, conseqüentemente cada amostra de entrada passou a ser representada por um vetor  $\bar{s}_i$ , isto é, para cada imagem existe agora um conjunto de características ao invés de apenas uma, sendo  $n$  o tamanho do conjunto de treinamento e  $k$  o número de funções de base radial usadas durante o treinamento.

Na equação (56), os elementos do vetor  $\bar{w}$  representam os pesos (o modelo) a serem encontrados. Cada um dos elementos do vetor  $\bar{f}_s$  é uma função de  $s$ , representando a média entre os limiares atribuídos por sujeitos humanos para a imagem, usando o experimento da Seção 3.2.

A matriz  $\Phi$  é a matriz de generalização, na qual cada elemento representa uma função de base radial  $\phi$  e  $\|\cdot\|$  representa uma norma euclidiana. Um exemplo do modelo usando um conjunto de 15 imagens para treinamento, 3 centros e como atributos a entropia e o desvio padrão de cada imagem pode ser observado na Figura 23. Nos gráficos a seguir, os atributos são representados pelos eixos  $x$  (desvio padrão),  $y$  (limiar) e  $z$  (entropia).



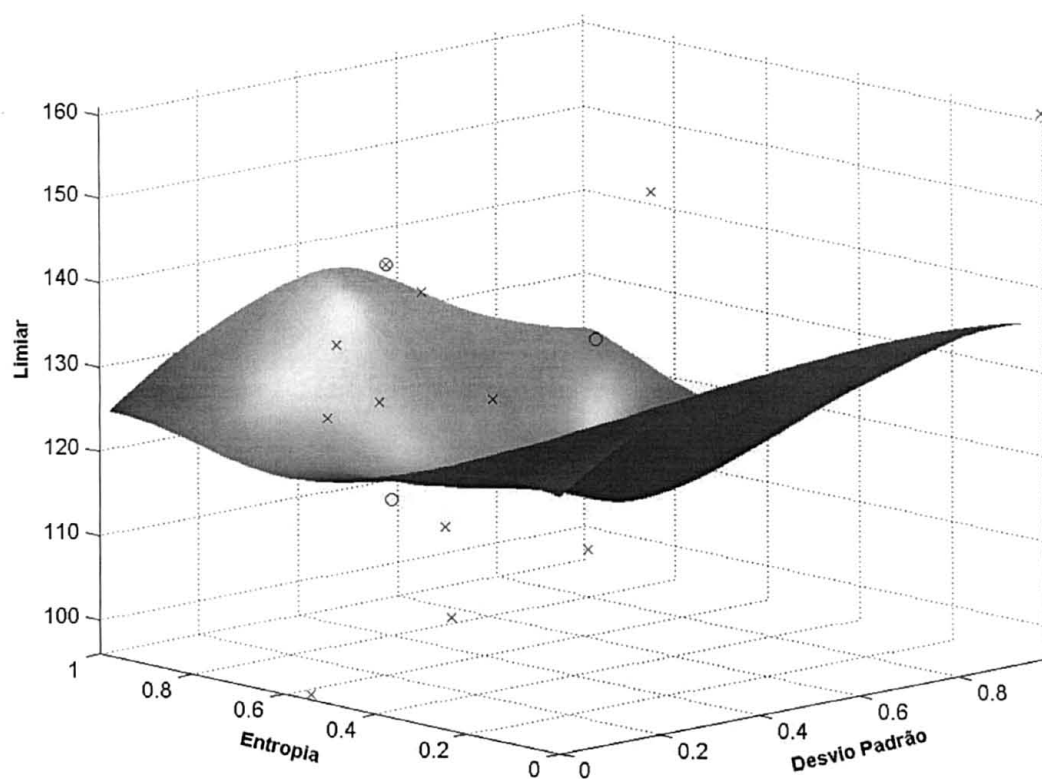


Figura 23 - Geração do modelo usando 3 centros. As cruzeiras representam as imagens usadas no treinamento e os círculos representam os centros selecionados pelo algoritmo k-médias.

Usando o mesmo conjunto de imagens pode-se observar na Figura 24 a geração do modelo usando 10 centros. Observamos que conforme o número de centros se aproxima da quantidade de dados disponíveis, a regressão fica mais rígida, podendo chegar à interpolação afim se for usado um número de centros igual ao número de imagens.

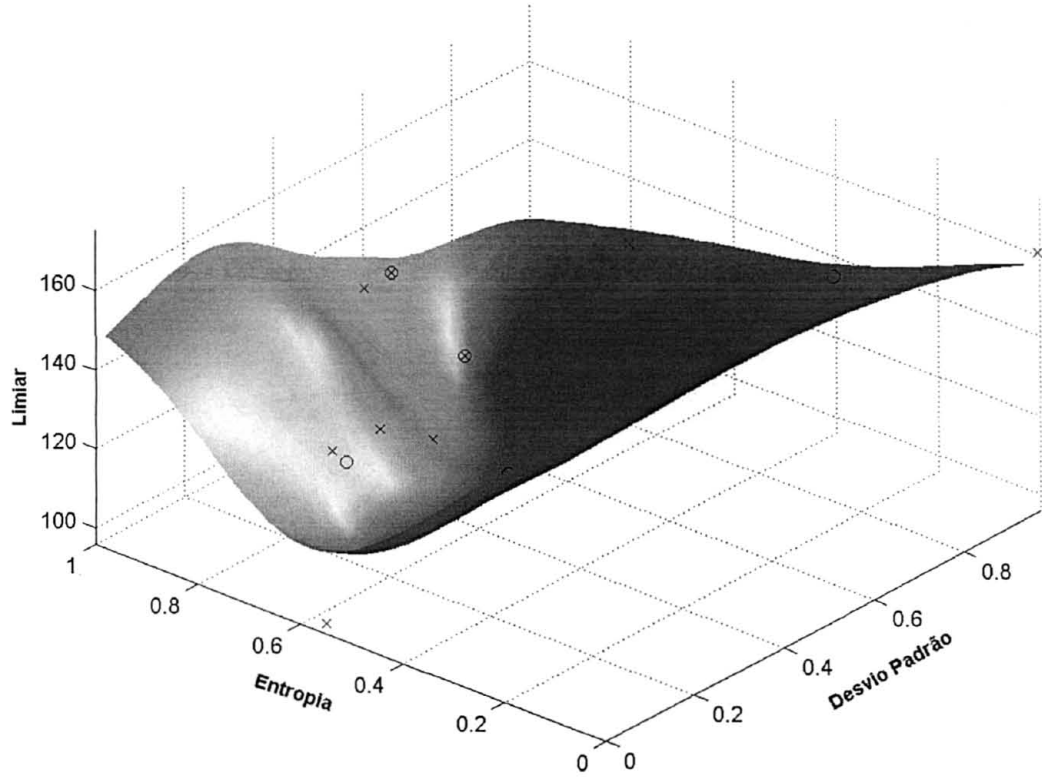


Figura 24- Geração do modelo usando 10 centros. As cruzeis representam as imagens usadas no treinamento e os círculos representam os centros selecionados pelo algoritmo k-médias.

Depois de encontrado o modelo, é usada a equação (57) para generalizar o conjunto de teste contendo as mesmas características usadas durante o treinamento, porém agora representadas por  $\bar{x}_j, j = 1, 2, \dots, m$ .

$$\begin{bmatrix} f(\bar{x}_1) \\ f(\bar{x}_2) \\ \vdots \\ f(\bar{x}_m) \end{bmatrix} = \begin{bmatrix} \phi_1(\|\bar{x}_1 - \bar{\mu}_1\|) & \phi_2(\|\bar{x}_1 - \bar{\mu}_2\|) & \cdots & \phi_k(\|\bar{x}_1 - \bar{\mu}_k\|) & 1 \\ \phi_1(\|\bar{x}_2 - \bar{\mu}_1\|) & \phi_2(\|\bar{x}_2 - \bar{\mu}_2\|) & \cdots & \phi_k(\|\bar{x}_2 - \bar{\mu}_k\|) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi_1(\|\bar{x}_m - \bar{\mu}_1\|) & \phi_2(\|\bar{x}_m - \bar{\mu}_2\|) & \cdots & \phi_k(\|\bar{x}_m - \bar{\mu}_k\|) & 1 \end{bmatrix}_{m \times k+1} \times \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \\ b \end{bmatrix} \quad (57)$$

Por fim, usando os pesos e centros encontrados durante o treinamento, as respostas desejadas são obtidas pela generalização dos dados de teste em função de seus dados de entrada, como pode ser observado na equação (58).

$$F(\bar{x}) = \sum_{i=1}^k w_i \phi(\|\bar{x} - \bar{\mu}_i\|) + b \quad (58)$$

Podemos observar a aplicação da equação (58), na Figura 25, para um conjunto de teste contendo 15 imagens e usando o modelo gerado com 3 centros.

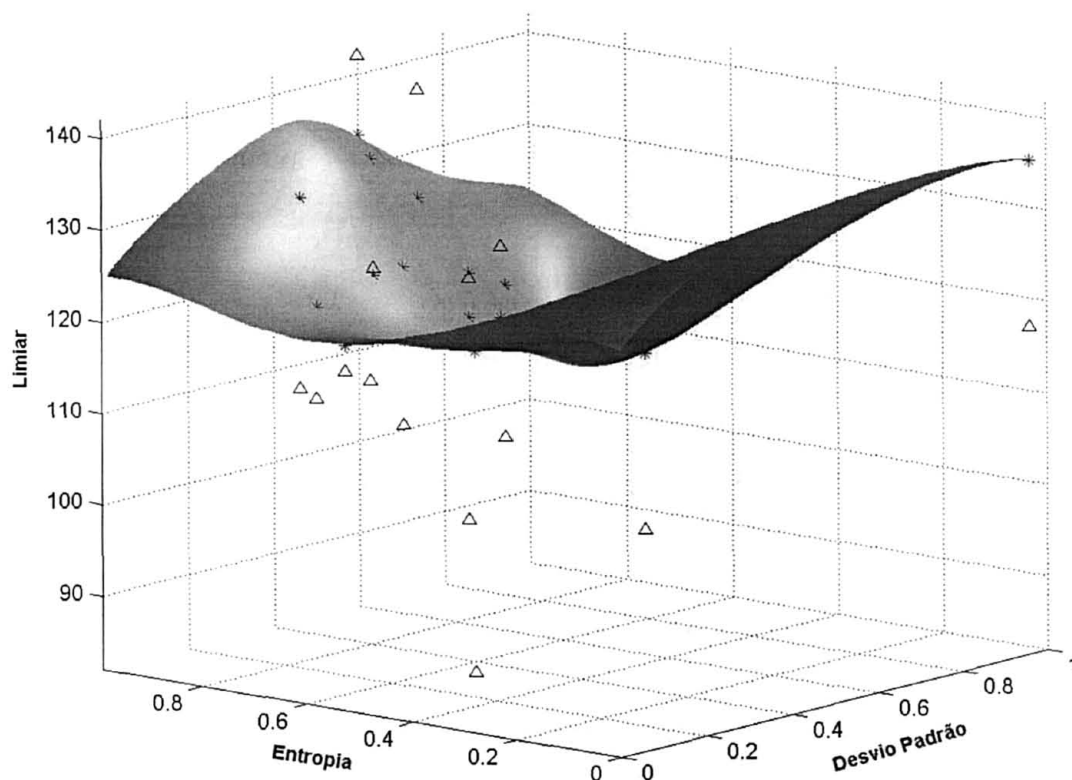


Figura 25- Aplicação do modelo usando 3 centros. Os triângulos representam as imagens de teste e os asteriscos representam as respostas do modelo para cada uma das imagens.

Na Figura 26 é ilustrada a aplicação do modelo para o mesmo conjunto de teste, mas agora usando a generalização com 10 centros.

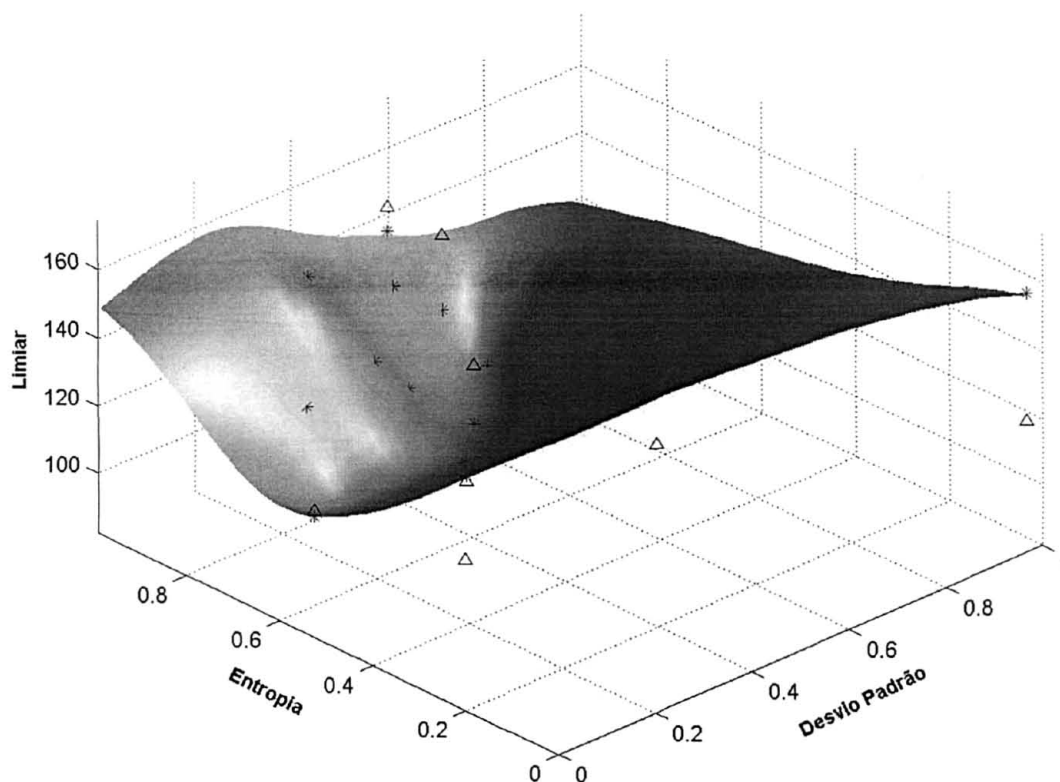


Figura 26- Aplicação do modelo usando 10 centros. Os triângulos representam as imagens de teste e os asteriscos representam as respostas do modelo para cada uma das imagens.

### 3.4. Validação da Aplicação do Modelo

O objetivo da validação do modelo é estimar os melhores valores para parâmetros como pesos, bias, atributos de entrada e quantidade de centros, para que realizem uma boa generalização para um conjunto de dados desconhecido, isto é, o conjunto de teste.

Mas, nem sempre é vantajoso utilizar, no modelo, as melhores combinações destes parâmetros. Por exemplo, ao aumentar o número de atributos e a quantidade de centros a complexidade também aumenta, e os resultados podem melhorar muito pouco ou até piorar muito [BISHOP, 1995].

A validação permite determinar o desempenho médio da rede, avaliando também se os conjuntos de treinamento e testes são adequados para atingir os objetivos definidos. No caso deste trabalho, determinar o limiar de acordo com a

percepção visual humana é o objetivo. No entanto, se os erros de validação forem muito grandes, é possível que as imagens fornecidas para teste e treinamento não tenham sido representativas.

A validação cruzada é uma ferramenta padrão para indicar a qualidade e a suavidade das respostas de um método de generalização de dados, como é o caso de uma rede de funções de base radial [HAYKIN, 2001].

Esta ferramenta tem base na divisão das amostras disponíveis em dois conjuntos: treinamento e teste. O treinamento é usado para determinar os parâmetros (pesos e bias) da rede, enquanto o teste é usado para estimar a qualidade (erro) proporcionada pelo treinamento [ORR, 1996].

Dividir as amostras em dois conjuntos, treinamento e teste, de uma forma específica pode influenciar nos resultados. A melhor decisão é dividir todo o conjunto de dados de diferentes formas e calcular a média dos erros apresentados durante as execuções dos treinamentos e testes, obtendo assim uma estimativa dos resultados reais do modelo.

Tendo em mente que diferentes métodos de validação oferecem análises diferentes é que foram utilizados os seguintes métodos de validação: generalizada, deixa um fora e bootstrap. Estes métodos foram utilizados sobre um conjunto de 110 imagens e com os atributos apresentados na seção 2.1. As validações testaram as respostas da rede começando com 1 centro, ou seja apenas uma função de base radial, e incrementou-se até 80 centros.

### **3.4.1. Validação Cruzada Generalizada**

A validação cruzada generalizada (VCG) conhecida também como  $k$ -validação cruzada, é uma das técnicas mais aplicadas na verificação da estabilidade de modelos [KOHAVI, 1995]. Seu funcionamento consiste em dividir o conjunto total de amostras em dois subconjuntos: treinamento e teste, mutuamente exclusivos como mostra a Figura 27.

O número  $k$  representa a quantidade de execuções e conseqüentemente o número de divisões das amostras em treinamento e teste, que serão usadas durante a validação, neste trabalho definido como 88, por representar 20% do total de imagens usadas na validação. A validação foi executada 88 vezes para cada quantidade de centros (variando de 1 até 80). Sendo o resultado de cada validação a média obtida durante as 88 execuções para cada número específico de centro.

Uma vantagem desse método é que todas as amostras são usadas para treinamento e teste, como pode ser observado na Figura 27. Na primeira execução o conjunto de teste é constituído pelas primeiras 22 amostras e o conjunto treinamento pelas restantes. Já na segunda execução, o conjunto de teste contém desde a 2ª até a 23ª amostra e o conjunto de treinamento com as restantes. Esse processo é repetido até que a validação tenha percorrido todas as amostras do conjunto.

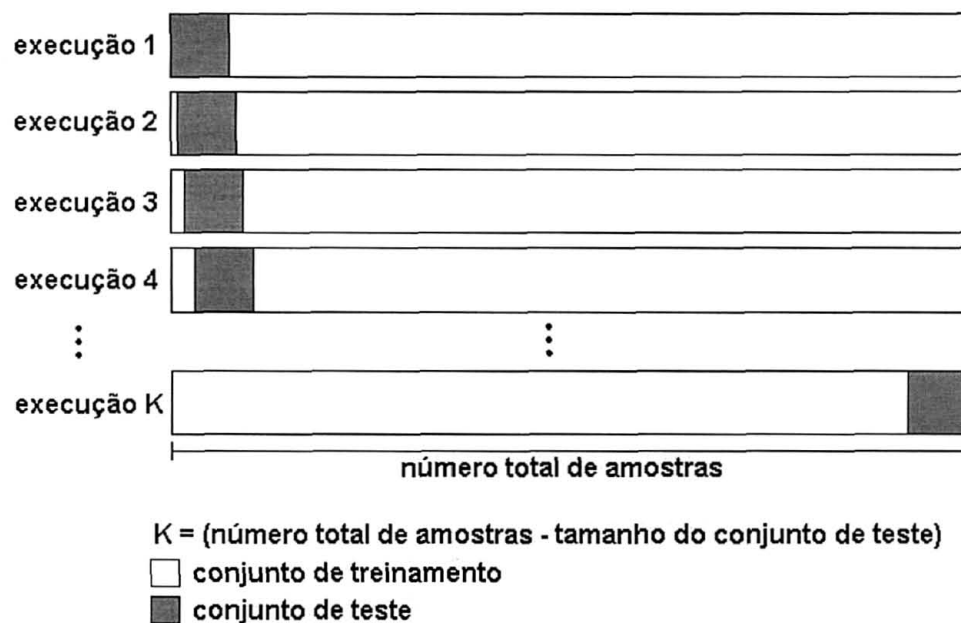


Figura 27 - Exemplo de divisão das amostras usadas pela validação cruzada generalizada <sup>9</sup>.

<sup>9</sup> OSUNA, R. G. Texas A&M University Computer Science Department, 2002. [http://faculty.cs.tamu.edu/rgutier/courses/cs790\\_wi02/l13.pdf](http://faculty.cs.tamu.edu/rgutier/courses/cs790_wi02/l13.pdf)

### 3.4.2. Validação Deixa-Um-Fora

A validação deixa-um-para (VDF) é uma forma extrema de validação cruzada. Essa validação atribui somente uma amostra ao conjunto de teste e todas as restantes são atribuídas ao conjunto de treinamento [BISHOP, 1995]. Assim, sua aplicação neste trabalho faz uso de 109 amostras para o treinamento e apenas uma para teste. Essa validação também pode ser vista como uma validação k-validação cruzada, sendo k definido como o tamanho do conjunto de amostras. A representação gráfica do funcionamento da validação deixa-um-para pode ser observada na Figura 28.

Dessa forma, essa validação fornece uma estimativa de erro para cada uma das amostras do conjunto.

Uma observação importante nessa validação é que todas as amostras disponíveis devem passar pelo menos uma vez pelo conjunto de teste, isto é, as amostras são combinadas para que todas sejam aproveitadas pelo treinamento e teste pelo menos uma vez.

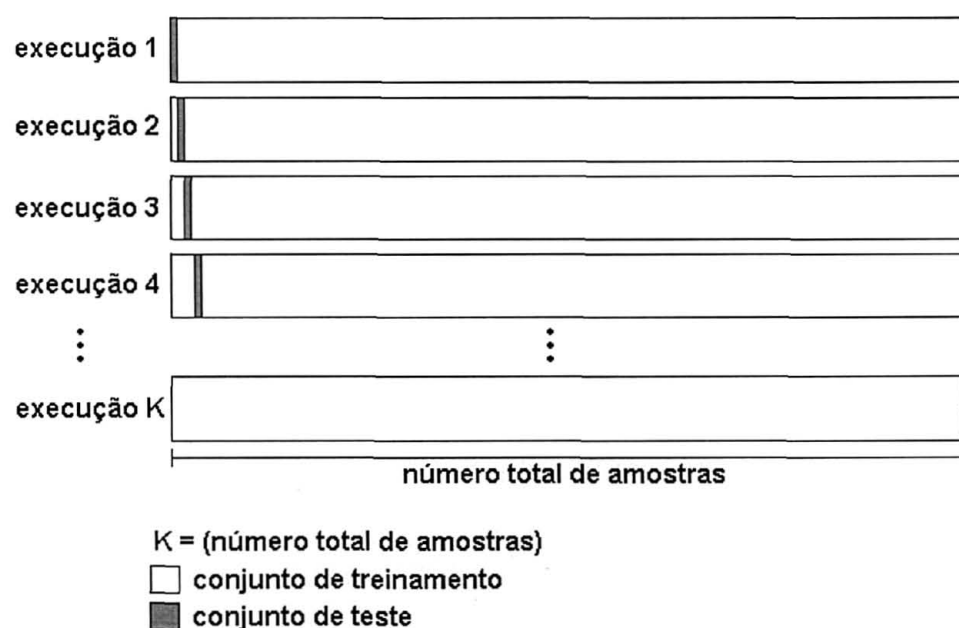


Figura 28 - Exemplo de divisão das amostras usadas pela validação deixa um para<sup>10</sup>.

<sup>10</sup> OSUNA, R. G. Texas A&M University Computer Science Department, 2002. [http://faculty.cs.tamu.edu/rgutier/courses/cs790\\_wi02/113.pdf](http://faculty.cs.tamu.edu/rgutier/courses/cs790_wi02/113.pdf)

Tendo dividido os conjuntos de treinamento e teste, a validação neste trabalho, é executada 110 vezes para cada quantidade de centros, começando em 1 até 80. O resultado da validação VDF é então obtido pela média dos erros quadráticos de todas as 110 execuções.

Segundo Teodoridis [THEODORIDIS & KONSTANTINOS, 1999], a maior desvantagem desse método é o tempo computacional usado para sua execução em relação à validação cruzada generalizada, e os resultados obtidos sobre a validação dos dados são superestimados, ou seja, fornece uma previsão pessimista sobre o conjunto de dados submetido à validação.

### **3.4.3. Validação Bootstrap**

A validação bootstrap foi proposta por Efron e Tibshirani [EFRON & TIBSHIRANI, 1993]. Esse método é a melhor ferramenta para estimar o desempenho de um conjunto de dados pequeno. Isso ocorre porque as  $N$  amostras de dados disponíveis são aleatoriamente atribuídas aos conjuntos de treinamento e teste durante uma execução, sem duplicação das amostras, ou seja, se uma amostra foi atribuída ao conjunto de treinamento, não poderá ser usada no teste. No entanto, neste método os dados poderão se repetir dentro de um conjunto, e conseqüentemente uma amostra poderá ser sorteada mais de uma vez para o mesmo conjunto.

Na execução seguinte todos os dados são novamente redistribuídos aleatoriamente aos conjuntos de treinamento e teste, havendo assim, sobreposição dos dados anteriormente usados, o que não acontece nas validações cruzadas.

Nessa validação foram desenvolvidas 40 execuções para cada número de centro, iniciando em 1 e incrementado até 80. O resultado é a média dos erros apresentados durante as execuções. O tamanho do conjunto de treinamento foi determinado com 88 amostras, restando 22 ao conjunto de teste.

A representação gráfica do desenvolvimento dessa validação pode ser observada na Figura 29.



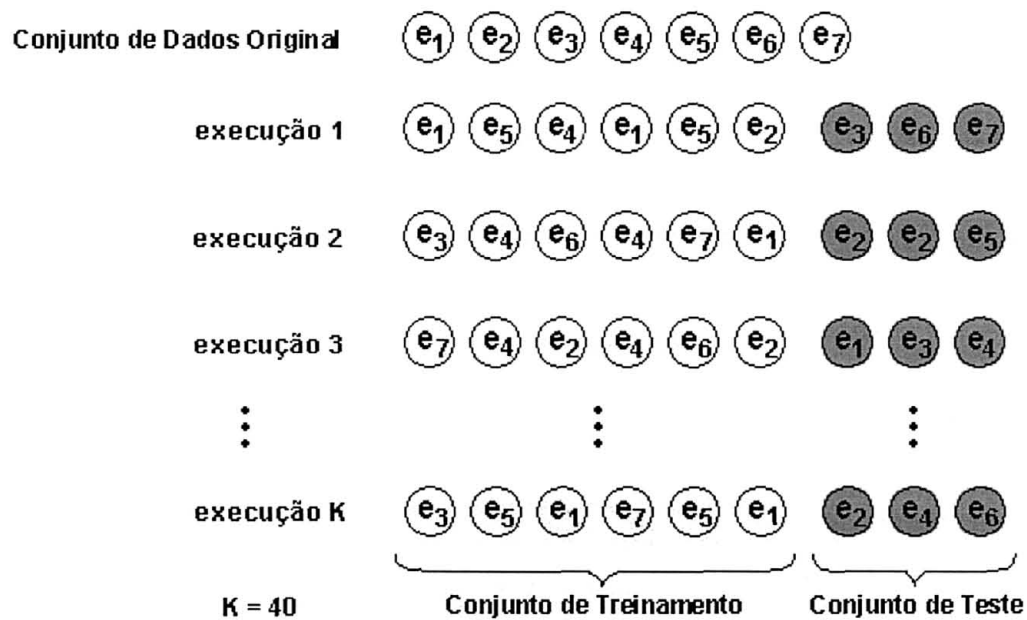


Figura 29- Exemplo de divisão das amostras usadas pela validação bootstrap<sup>11</sup>.

### 3.5. Comparação entre Modelos de Limiarização

Uma vez conhecida a percepção humana (limiares) para um conjunto de 110 imagens e os limiares automáticos propostos por Otsu [OTSU, 1978], Kapur [KAPUR et. al., 1985] e Brink [BRINK, 1996], tem início a análise de semelhanças entre os limiares humanos e automáticos, a qual é um dos objetivos desse trabalho.

Para esta tarefa é usado o coeficiente de correlação entre os limiares humanos e os automáticos. A correlação ( $r$ ) é uma medida do grau de associação entre duas características indicando o quanto uma característica contribui para a predição da outra. Seu valor segue uma escala:  $-1 \leq r \leq 1$ . Quando  $r = 1$ , significa que existe um grau máximo de associação, isto é, a ocorrência de uma característica é diretamente proporcional a ocorrência da outra característica, quando  $r = -1$  existe um comportamento contrário (inversamente proporcional) e se  $r = 0$  significa que a ocorrência de uma característica em nada contribui para a ocorrência da outra, isto é, elas são independentes.

<sup>11</sup> OSUNA, R. G. Texas A&M University Computer Science Department, 2002. [http://faculty.cs.tamu.edu/rgutier/courses/cs790\\_wi02/113.pdf](http://faculty.cs.tamu.edu/rgutier/courses/cs790_wi02/113.pdf)

O coeficiente de correlação entre duas características:  $x$  e  $y$ , é dado pela equação:

$$R = \frac{\sum (x - \bar{x}) \times (y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \times \sum (y - \bar{y})^2}} \quad (59)$$

Sendo:

$\bar{x}$ , a média da característica  $x$ .

$\bar{y}$ , a média da característica  $y$ .

Assim, para a determinação do coeficiente de correlação, foi desenvolvido um aplicativo usando a linguagem Delphi versão 6, o qual plota um diagrama de dispersão entre limiares humanos x limiares automáticos, o qual segundo [MOORE, 2000] é a maneira mais eficiente de mostrar a relação entre duas variáveis quantitativas.

Simultaneamente à plotagem do diagrama, o aplicativo também determina o coeficiente de correlação entre os dados. Após a determinação da correlação também é aplicado aos dados um ajuste de curvas usando o método dos mínimos quadrados, o qual define a reta que melhor ajusta o conjunto de dados no sentido de minimizar a soma dos quadrados das distâncias dos pontos à reta [MOORE, 2000]. Uma captura de tela desse aplicativo pode ser vista na Figura 30.

Observando a Figura 30 percebe-se que os dados e também a reta que melhor ajusta os dados estão plotados no detalhe 5 (diagrama de dispersão). Os dados usados foram selecionados de acordo com a caixa de seleção no detalhe 1, assim o usuário pode escolher qual conjunto de dados irá usar. No detalhe 2, o usuário pode escolher qual será o limiar considerado na comparação com o limiar humano, sendo os seis primeiros itens usados pelo modelo RFBR. Na primeira linha do detalhe 3, se tem a correlação entre os limiares considerando todo o conjunto de dados. No detalhe 4, é disponibilizado ao usuário a possibilidade de remover os pontos discrepantes (*outliers*), por meio de sua distância à reta de ajuste. Dessa forma, o usuário pode escolher o limiar para desconsiderar os pontos na determinação da correlação, e assim obter uma correlação mais realista do conjunto de dados, o qual é exibido na segunda linha do detalhe 3, e assim, os limiares com maior distância à reta de ajuste que o limiar escolhido são identificados, sendo plotados com cores diferentes. E para o ponto de

distância máxima é traçada a reta de sua distância e sua identificação é diferenciada dos demais pontos que ficaram de fora do cálculo da correlação.

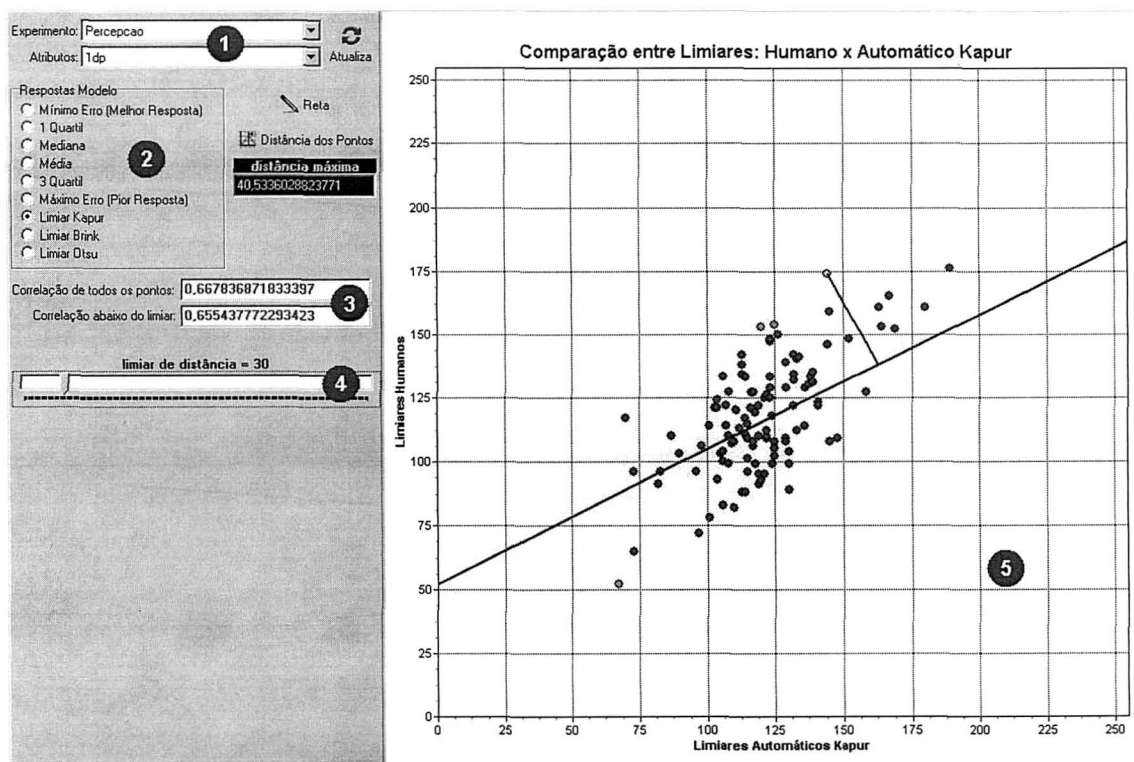


Figura 30- Software desenvolvido para comparação entre limiares.

## 4 RESULTADOS

Este capítulo apresenta inicialmente os resultados da aplicação do modelo de limiarização RFBR (seção 3.3), comportando as três primeiras seções. Cada seção representa uma etapa de evolução do modelo, e em cada etapa são apresentados os resultados obtidos.

A primeira seção apresenta os resultados alcançados com o modelo inicial (TPS), os quais motivaram a continuidade e aperfeiçoamento do modelo RFBR. A segunda seção apresenta os resultados da fase intermediária do modelo e finalmente na terceira seção os resultados obtidos com o modelo final da RFBR.

Nas duas primeiras fases de desenvolvimento foi usado um conjunto contendo 110 imagens variadas (pessoas, paisagens, carros e casas) e com tamanhos diferentes. Foram extraídos seus respectivos atributos: entropia, intensidade média, desvio padrão, e medidas de contraste, os quais estão descritos na seção 2.1. Os limiares atribuídos a esse conjunto foram extraídos da percepção de apenas um sujeito, utilizando o experimento psicofísico apresentado na seção 3.2.

Na fase final foi usado este mesmo conjunto de 110 imagens, mas agora com a média dos limiares humanos atribuídos pelas percepções coletadas durante a execução do experimento psicofísico com diversos sujeitos, conforme descrito na seção 3.2. Nesta fase também são mostrados os resultados alcançados com a aplicação do modelo em um conjunto de imagens contendo apenas faces de pessoas.

No apêndice podem ser observados a descrição das imagens, o valor de cada um dos atributos e seus respectivos limiares automáticos e humanos para o conjunto de imagens variadas e também para o conjunto de imagens de faces.

Em todas as três fases, os atributos das imagens foram normalizados, com os valores dos atributos variando dentro do intervalo  $[0,1]$ .

Nas duas primeiras fases de desenvolvimento, as validações consistiram de alguns testes simples para atestar a qualidade do modelo. Já na terceira e última fase (seção 4.3), métodos de validação reconhecidos foram aplicados.

A seção 4.4 apresenta os resultados de uma comparação entre a percepção visual humana e alguns métodos de limiarização automáticos, observando as correlações encontradas. Esta comparação continua na seção 4.5, mas agora usando os resultados do modelo RFBR para compará-los aos dados coletados no experimento psicofísico. Novamente são apontadas as correlações encontradas durante a análise.

#### **4.1. Resultados Iniciais**

Nesta seção são apresentados os resultados obtidos com o modelo de interpolação usando thin-plate splines (TPS), os quais representam o ponto de partida para o desenvolvimento e aperfeiçoamento desse modelo.

Na interpolação, o conjunto contendo 110 imagens variadas e um único atributo, a entropia, foi dividido aleatoriamente em 16 experimentos com tamanhos variados. A única restrição observada na divisão foi que as imagens com entropias extremas (máximas e mínimas) foram mantidas no conjunto de treinamento. Esse cuidado teve o objetivo de garantir que as amostras do conjunto de teste estivessem dentro do intervalo de treinamento. Assim, cada um dos experimentos foi então dividido em treinamento e teste. A divisão dos experimentos, bem como seus respectivos resultados são mostrados na Tabela 3.

Essas divisões das imagens em 16 experimentos e seus respectivos testes foram realizadas como uma forma de validar esse modelo (interpolação). A medida de desempenho foi o erro médio quadrático entre as respostas do modelo e as respostas humanas ao experimento (corretas), bem como o desvio padrão destes erros.

Como pode ser observado na Tabela 3, os experimentos 15 e 16 influenciaram muito para o aumento da média do erro e desvio padrão. Isto ocorreu porque as imagens estavam muito próximas umas das outras, provocando a instabilidade da matriz de interpolação. Essa instabilidade ocasionou oscilações indesejáveis na interpolação como mostra a Figura 32. Sem os resultados dos experimentos 15 e 16, a média do erro RMS foi 8,65 e a média do desvio padrão foi 19,03.

Número do Experimento	Imagens Treinamento	Imagens Teste	Erro RMS	Desvio Padrão
1	11	11	10,2221	18,3087
2	6	6	8,4877	11,6792
3	10	10	7,9373	12,5173
4	15	15	3,4817	8,0945
5	15	16	7,1865	22,6843
6	15	15	5,5318	11,8412
7	18	18	16,2378	46,627
8	9	10	8,6175	12,6696
9	9	9	5,2508	9,7044
10	14	17	4,3257	13,7727
11	13	13	9,8166	20,1540
12	10	10	16,7383	34,7534
13	18	18	11,3613	31,9238
14	15	15	5,9133	11,7932
15	88	22	23,0060	78,2444
16	62	47	58,7119	259,8208
<b>Média</b>			<b>12,6766</b>	<b>37,7867</b>

Tabela 3 - Resultados da validação do modelo TPS feito com alguns testes simples, usando como atributo apenas a entropia das imagens.

O experimento que apresentou o melhor desempenho foi o número 4, o qual pode ser visto na Figura 31, seguido do experimento 16 com o pior resultado, apresentado na Figura 32.

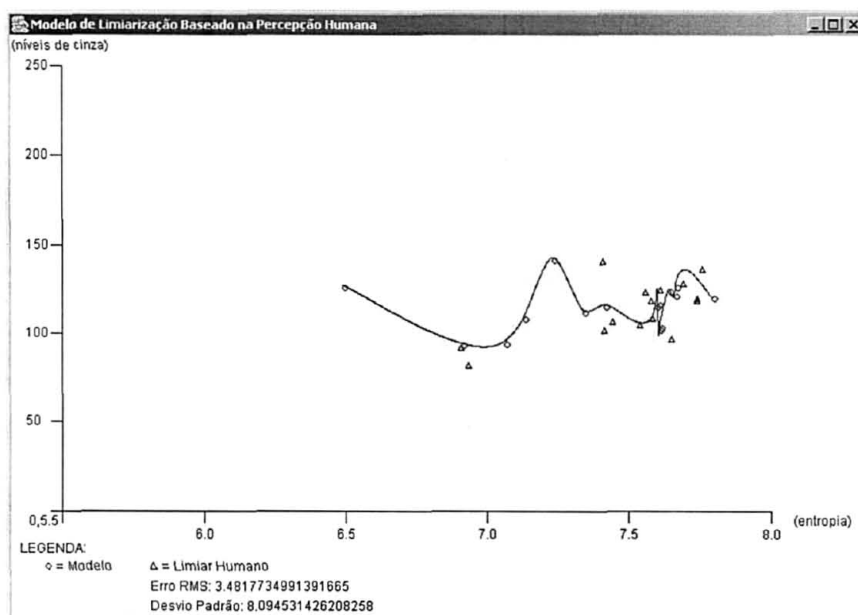


Figura 31 - Experimento com melhor desempenho na fase inicial.

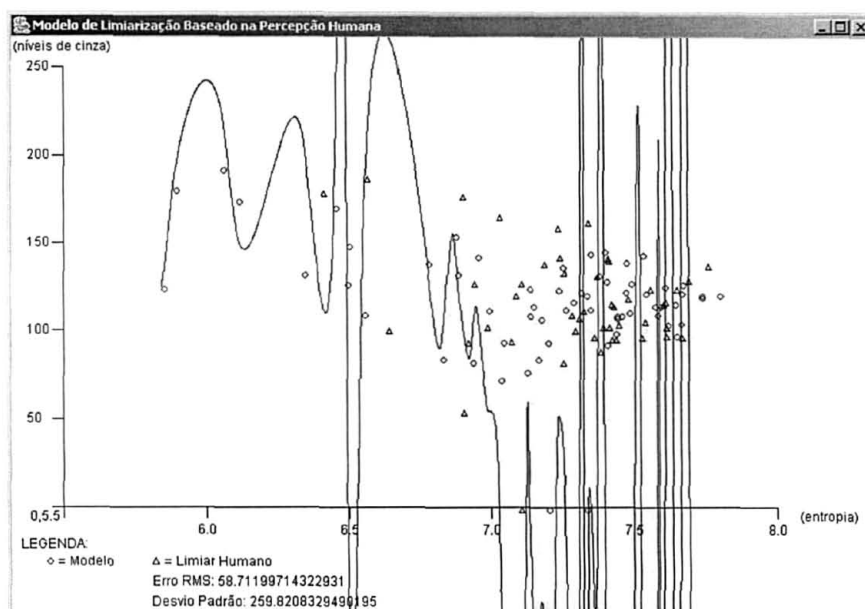


Figura 32 - Experimento com pior desempenho na fase inicial.

O modelo de interpolação usando TPS também foi testado com dados gerados a partir de funções com comportamentos conhecidos como a linear, a logarítmica, a exponencial, a polinomial e um padrão aleatório.

Para cada comportamento, gerou-se um conjunto de dados contendo 100 elementos. Em cada conjunto, os 100 elementos foram divididos aleatoriamente em 2 subconjuntos: treinamento e teste, cada um com 50 valores. A Tabela 4 apresenta os resultados médios obtidos para 10 execuções de cada conjunto.

Nome do Experimento	Amostras de Treinamento	Amostras de Teste	Erro RMS médio	Desvio Padrão Médio
Aleatório	50	50	184,2201	978,3368
Exponencial	50	50	0,6111	3,4635
Linear	50	50	0,2720	1,4161
Logarítmico	50	50	0,3023	1,6728
Polinomial	50	50	0,5724	3,1489

Tabela 4 - Resultados da validação unidimensional (entropia), que representa a média entre 10 execuções do modelo para os comportamentos conhecidos.

A Tabela 4 mostra que, para o comportamento aleatório, o resultado apresentou um erro muito grande, como já era de se esperar. Por outro lado, para todos

os outros comportamentos o modelo se mostrou bastante eficiente quanto as suas generalizações.

#### 4.2. Resultados intermediários

Motivados pelos resultados iniciais, passou-se a considerar dois atributos no modelo. Como consequência desse acréscimo foram refeitos os testes iniciais para avaliar a inclusão deste novo atributo.

Os testes iniciaram com os mesmos comportamentos conhecidos usados anteriormente, mas agora cada função recebeu dois parâmetros, retornando um valor de acordo com o comportamento pré-estabelecido, gerando uma curva entre os dois parâmetros.

Assim, foi gerado novamente, para cada um dos comportamentos conhecidos, um conjunto de dados contendo 100 elementos. Dentro de cada conjunto, os 100 elementos foram divididos aleatoriamente em 2 subconjuntos: treinamento e teste, cada um com 50 valores. A Tabela 5 apresenta os resultados médios obtidos por 10 execuções (treinamentos e testes) para cada conjunto.

Como pode ser observado na Tabela 5, os resultados obtidos com a inclusão do novo atributo não melhorou o modelo, uma vez que o erro RMS e o desvio padrão aumentaram. O único resultado positivo foi conseguido com o comportamento linear, que apresentou um erro RMS de 0,6589 e desvio padrão de 2,3404, mas mesmo assim não foi melhor que o resultado unidimensional.

Nome do Experimento	Amostras de Treinamento	Amostras de Teste	Erro RMS médio	Desvio Padrão Médio
Aleatório	50	50	152,058	790,8187
Exponencial	50	50	15,831	84,348
Linear	50	50	0,6589	3,6289
Logarítmico	50	50	2,1342	11,7362
Polinomial	50	50	95,937	607,979

Tabela 5 - Resultados da validação bidimensional (entropia e variância), que representa a média entre 10 execuções do modelo para os comportamentos conhecidos.



Como os resultados da interpolação de dados com dois atributos e com comportamentos conhecidos não foram satisfatórios, concluiu-se que também não seria adequado como um modelo de limiarização, uma vez que no caso unidimensional os comportamentos conhecidos tiveram um bom desempenho. Mesmo assim, os testes com as imagens, representadas por atributos de entropia e variância, foram realizados e os resultados confirmaram a suspeita inicial, isto é, não foram satisfatórios, como pode ser observado na Tabela 6.

Número do Experimento	Imagens Treinamento	Imagens Teste	Erro RMS	Desvio Padrão
1	11	11	10,0353	21,9674
2	6	6	8,0566	7,46118
3	10	10	38,4595	98,675
4	15	15	13,4169	27,4688
5	15	16	7,4962	18,494
6	15	15	6,6746	14,9744
7	18	18	38,9169	117,7792
8	9	10	15,1034	34,8123
9	9	9	57,7048	131,1379
10	14	17	14,2044	26,6145
11	13	13	75,7903	218,7371
12	10	10	150,8284	338,2793
13	18	18	13,3030	40,2505
14	15	15	76,8462	243,6692
15	88	22	288,0557	1125,0
16	62	47	66,5344	342,1789
<b>Média</b>			<b>55,0891</b>	<b>175,4687</b>

Tabela 6 - Resultados da validação do modelo TPS feito com alguns testes simples, usando como atributos a entropia e desvio padrão das imagens.

Estes resultados com imagens levantaram a suspeita de que a interpolação com TPS poderia não ser a maneira mais adequada de generalizar as respostas perceptivas, e não que a variância fosse um atributo irrelevante. Outra suspeita caiu sobre a interpolação, por obrigar que a função ajustada mapeie todos os pontos utilizados, tornando-a uma maneira rígida de generalizar.

Na tentativa de superar essa limitação, ou seja, a rigidez da generalização, é que se adotou um modelo baseado em Rede de Funções de Base Radial (RFBR).

### 4.3. Resultados Finais

Foram testados inicialmente os mesmos atributos, entropia e variância, para os mesmos conjuntos de dados, ou seja, as 110 imagens variadas e os comportamentos conhecidos.

Também foi gerado para cada um dos comportamentos conhecidos, além de um conjunto com 100 pares aleatórios de atributos, um novo conjunto de dados contendo 50 pares, também aleatórios, de atributos. Cada um dos conjuntos foi dividido aleatoriamente em dois outros conjuntos, treinamento e teste, com o mesmo número de elementos.

O teste de desempenho do modelo RFBR na aplicação aos conjuntos citados acima consistiu em três tipos de validação: validação cruzada generalizada (VCG), validação deixa-um-fora (VDF) e *bootstrap* (seção 3.4), como uma forma de apresentar resultados mais consistentes.

Um parâmetro fundamental em generalizações usando RFBRs é o número de centros, e conseqüentemente de funções, usados durante o treinamento, ou seja, na construção do modelo. Estes centros são os mesmos usados na generalização (teste). A Tabela 7 mostra os resultados obtidos na generalização dos comportamentos conhecidos, bem como as quantidades de centros (coluna FBR) usados para determiná-las.

A Tabela 7 chama a atenção por mostrar que quase todos os resultados ficaram próximos aos resultados obtidos com a interpolação TPS unidimensional. Além disso, a RFBR se mostrou computacionalmente mais eficiente, pois usou uma menor quantidade de centros (ou funções) e conseqüentemente sistemas lineares de menor ordem. Apresentou ainda resultados mais robustos, por permitir validações mais confiáveis.

Nome do Experimento	Amostras de Treinamento	Amostras de Teste	VDF		VCG		Bootstrap	
			Erro RMS	FBR	Erro RMS	FBR	Erro RMS	FBR
Aleatório	80	20	57,2880	3	15,5141	1	15,7971	3
Aleatório	40	10	55,7851	1	20,2679	1	20,7675	1
Exponencial	80	20	2,6669	74	2,0600	59	1,5158	52
Exponencial	40	10	4,3016	40	5,0765	31	2,8437	35
Linear	80	20	0,7015	80	0,7561	48	0,3559	60
Linear	40	10	1,9872	40	2,7731	34	1,3124	30
Logarítmico	80	20	0,6525	72	1,0146	32	0,2763	57
Logarítmico	40	10	0,8087	40	1,0338	29	0,4269	32
Polinomial	80	20	1,1938	63	1,9721	64	0,7648	60
Polinomial	40	10	2,1190	39	2,3852	32	1,6580	30

Tabela 7 - Resultados da validação bidimensional (entropia e variância) do modelo para os comportamentos conhecidos. Os erros RMS representam a média entre 96 execuções para a VCG e VDF e 40 execuções para o bootstrap.

Quando usadas amostras com 50 elementos, a RFBR encontrou os melhores resultados com 40 centros, isto é, uma superfície com pouca suavidade. Por outro lado, para dados aleatórios, a rede encontrou melhores resultados usando apenas 1 centro, ou seja, a máxima suavidade. Usando amostras com 100 elementos, observou-se resultados similares, mas com erros menores, sugerindo que aumentando a quantidade de amostras (treinamento) o modelo RFBR tende a errar menos e conseqüentemente apresentar melhor desempenho.

Os resultados com dados obtidos de funções de controle simulando um comportamento conhecido motivaram a aplicação do modelo em um conjunto de imagens variadas usando os atributos entropia e variância. Contudo, desta vez aplicou-se o método da generalização por RFBRs a todo o conjunto, já que as validações permitiram separações entre conjuntos de treinamento e teste de forma aleatória.

O melhor resultado obtido com a validação aplicada ao conjunto de imagens variadas pode ser observado na Tabela 8, notando-se que os erros variam de acordo com o método de validação. Isso ocorre porque cada método de validação utiliza uma estratégia diferente na escolha da divisão dos subconjuntos de treinamento e teste. Observa-se também que, comparados aos resultados obtidos pela interpolação com TPSs, a aplicação das RFBRs resultaram em erros menores em dois métodos de validação. Isto sugere que a capacidade de generalização das RFBRs pode ser melhor

explorada na busca de atributos representativos na implementação de modelos de percepção. O número de centros usados foi pequeno (11 e 13), indicando que a superfície que generaliza os atributos entropia e desvio padrão é suave.

Nome do Experimento	Imagens Treinamento	Imagens Teste	VDF		VCG		Bootstrap	
			Erro RMS	FBR	Erro RMS	FBR	Erro RMS	FBR
Imagens variadas	88	22	16,6143	13	4,4469	11	4,3226	13

Tabela 8 - Resultados da validação bidimensional (entropia e variância) do modelo para os comportamentos conhecidos. Os erros RMS representam a média entre 88 execuções para a VCG e VDF e 40 execuções para o bootstrap.

A Figura 33 ilustra a superfície gerada (modelo), representando graficamente os resultados da Tabela 8. Os círculos representam os centros e os triângulos as imagens usadas durante o teste (validação).

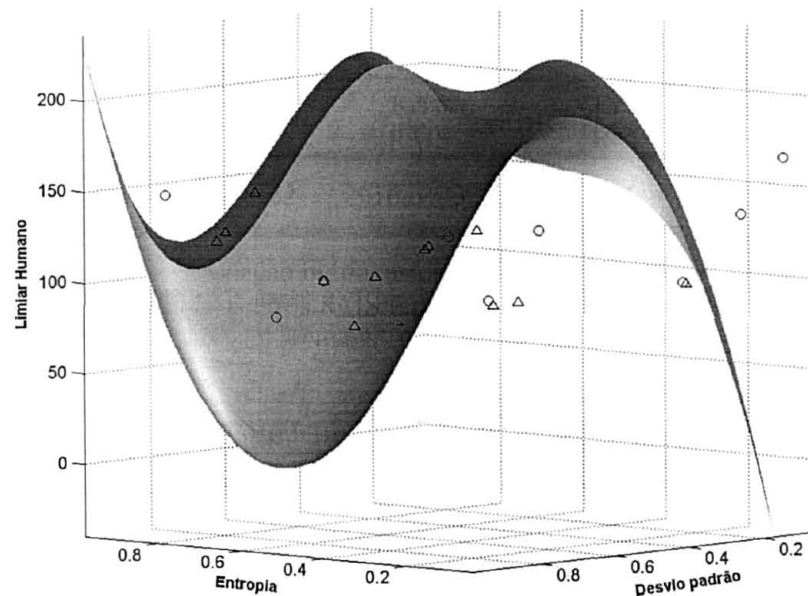


Figura 33 - Modelo de limiarização usando os atributos: entropia, variância e os limiares humanos com 13 funções de base radial (FBR).

Procurando explorar a capacidade da abordagem em selecionar características, experimentou-se a inclusão de novos atributos das imagens aplicados ao modelo de limiarização baseado nas respostas dos sujeitos humanos. Tomando como referência as validações, poderiam ser estimados: *i)* o número de centros apropriados; *ii)* quantos atributos seriam suficientes para representar as respostas humanas; e *iii)* quais

atributos seriam os mais representativos. Dessa forma, foi testado, para o conjunto de imagens variadas, todas as combinações de atributos e os resultados obtidos são apresentados na Tabela 9. Nas validações VCG e *bootstrap*, o conjunto foi dividido em treinamento com 88 imagens e teste com 22, aleatoriamente.

Atributos	VDF		VCG		Bootstrap	
	Erro RMS	FBR	Erro RMS	FBR	Erro RMS	FBR
1dp	17,6998	10	4,5138	1	4,5985	3
1e	16,2575	2	4,3868	3	4,2184	4
1im	14,0309	9	3,8271	4	3,6413	5
1mev4	18,2888	1	4,5427	1	4,7695	1
2dp-im	13,5481	41	3,8213	34	3,7629	29
2dp-maxv4	17,4628	35	4,5806	1	4,5052	12
2dp-mev4	17,1863	13	4,5868	1	4,7683	1
2e-dp	16,6143	13	4,4469	11	4,3226	13
2e-im	14,3666	24	4,1264	5	3,9587	5
2e-maxv4	16,8571	16	4,5877	1	4,5608	7
2e-mev4	16,9976	9	4,5985	1	4,4642	11
2im-maxv4	15,2197	19	4,3124	4	4,2249	21
2im-mev4	15,2532	25	4,2669	13	4,1669	10
3dp-im-maxv4	13,9894	46	3,9011	37	3,9391	41
3dp-im-mev4	13,5906	36	3,8804	31	3,8296	30
3dp-qlv4-q3v4	17,4616	14	4,5490	1	4,6052	12
3e-dp-im	14,0887	27	4,1854	20	3,7988	21
3e-dp-maxv4	17,4299	33	4,5637	1	4,4904	21
3e-dp-mev4	16,3363	19	4,5691	1	4,4496	20
3e-im-maxv4	15,3454	32	4,2549	11	4,1101	27
3e-im-mev4	14,9429	13	4,2027	14	3,9686	13
3e-qlv4-q3v4	15,7676	18	4,5009	2	4,4244	15
3im-qlv4-q3v4	13,5641	38	3,8784	41	3,7673	33
4dp-im-qlv4-q3v4	13,1500	76	3,8813	48	3,6658	46
4e-dp-im-maxv4	14,2882	20	4,2464	22	3,9809	25
4e-dp-im-mev4	13,5751	26	4,3095	25	3,8817	38
4e-dp-qlv4-q3v4	16,3416	19	4,5198	1	4,3490	21
4e-im-qlv4-q3v4	14,1323	42	3,8802	40	3,9245	39
5e-dp-im-mev4-maxv4	14,4605	30	4,3988	24	4,0399	29
6e-dp-im-mev4-qlv4-q3v4	13,5199	37	4,1734	39	3,8561	36
7e-dp-im-mev4-maxv4-qlv4-q3v4	13,9582	31	4,2000	33	3,8278	38

Tabela 9 - Resultados da validação aplicada a vários atributos e suas combinações ao modelo para o conjunto de imagens variadas. Os erros RMS representam a média entre 88 execuções para a VCG e VDF e 40 execuções para o bootstrap. Os atributos estão codificados, iniciando pelo número que significa a quantidade de atributos combinados e seguido das seguintes abreviações: e = entropia, dp = desvio padrão, im = intensidade média, mev4 = mediana do histograma de contraste, qlv4 = primeiro quartil do histograma de contraste, q3v4 = terceiro quartil do histograma de contraste e maxv4 = máximo do histograma de contraste.

É de se destacar que usando apenas um atributo, a intensidade média, o erro foi o menor e com apenas 6 centros (em média). Por outro lado, a mediana do histograma de contraste ficou com o pior resultado mesmo usando a suavidade máxima do modelo (1 centro).

Usando uma combinação de dois atributos, aquela que resultou no menor erro foi a combinação entre desvio padrão e intensidade média, usando uma suavização intermediária. O maior erro foi da combinação entre desvio padrão e o máximo do histograma de contraste, mesmo usando um nível maior de suavidade.

Com três atributos, o melhor desempenho foi com a combinação entre desvio padrão, intensidade média e mediana, com um nível intermediário de suavização. O pior resultado foi usando a combinação entre desvio padrão, primeiro quartil e terceiro quartil.

A combinação de quatro atributos foi aquela que indicou o menor erro e foi obtida com a combinação entre desvio padrão, intensidade média, primeiro quartil e terceiro quartil do histograma de contraste.

É possível perceber, nestes resultados, um padrão na escolha do número de centros usados pelos experimentos que apresentaram os piores resultados. Os piores resultados usam uma quantidade relativamente pequena de centros na validação deixo-um-fora, uma quantidade intermediária na validação bootstrap e sempre um centro na validação cruzada generalizada.

A abordagem da RFBR foi aplicada também em um conjunto de imagens menos diversas. Um conjunto de 59 imagens de faces foi utilizado, todas com o mesmo tamanho (71 x 92 pixels) e com faces na mesma posição. As condições de captura da imagem não foram controladas e um exemplo típico é mostrado na Figura 34.



Figura 34 - Exemplo de imagem usada no conjunto de faces.

O conjunto de 59 imagens foi dividido aleatoriamente entre um conjunto de treinamento, com 48 imagens, e outro de teste, com 11 imagens para as validações VCG e *bootstrap*. Os resultados da aplicação da generalização podem ser observados na Tabela 10.

Atributos	VDF		VCG		Bootstrap	
	Erro RMS	FBR	Erro RMS	FBR	Erro RMS	FBR
1dp	17,6196	2	6,4493	1	5,9856	1
1e	17,7627	1	6,7145	3	6,1959	4
1im	15,1456	9	5,8069	2	5,5498	7
1mev4	17,7879	6	6,4312	1	6,4458	2
2dp-im	16,2182	4	6,1539	3	6,3400	1
2dp-maxv4	17,6277	1	6,5290	1	6,1892	1
2dp-mev4	18,4177	1	6,7625	1	6,5659	1
2e-dp	15,9965	21	6,5928	1	6,2932	14
2e-im	15,4314	15	6,2492	2	6,0951	7
2e-maxv4	17,8206	10	6,4639	1	6,3332	3
2e-mev4	15,5360	26	6,5387	1	6,4157	1
2im-maxv4	16,3235	2	5,9648	3	5,8886	3
2im-mev4	14,0864	30	6,1835	15	6,0743	26
3dp-im-maxv4	16,0457	7	6,1712	7	5,9863	2
3dp-im-mev4	15,0996	22	6,1715	21	6,3236	4
3dp-qlv4-q3v4	18,2864	6	6,7587	1	6,4204	6
3e-dp-im	15,5346	18	6,3158	9	6,2546	5
3e-dp-maxv4	17,1145	5	6,4296	6	6,2593	5
3e-dp-mev4	16,8471	7	6,3329	6	6,4130	7
3e-im-maxv4	16,1942	38	6,3320	1	6,1069	6
3e-im-mev4	16,1535	27	6,4086	4	6,1555	3
3e-qlv4-q3v4	15,8826	19	6,6275	10	6,1134	17
3im-qlv4-q3v4	15,9100	12	6,3407	4	5,9438	5
4dp-im-qlv4-q3v4	16,7338	6	6,3951	2	6,0968	4
4e-dp-im-maxv4	15,4068	10	6,2475	10	5,9698	2
4e-dp-im-mev4	15,2988	25	6,2628	14	6,1753	16
4e-dp-qlv4-q3v4	16,0006	11	6,4438	6	6,0442	6
4e-im-qlv4-q3v4	15,6502	32	6,3124	3	6,0325	9
5e-dp-im-mev4-maxv4	15,9626	24	6,3781	13	6,4578	2
6e-dp-im-mev4-qlv4-q3v4	16,4635	14	6,3170	1	6,1624	1
7e-dp-im-mev4-maxv4-qlv4-q3v4	15,8824	12	6,0707	1	5,9791	1

Tabela 10 - Resultados da validação aplicada a vários atributos e suas combinações ao modelo para o conjunto de imagens de faces. Os erros RMS representam a média entre 45 execuções para a VCG e VDF e 40 execuções para o bootstrap. Os atributos estão codificados, iniciando pelo número que significa a quantidade de atributos combinados e seguido das seguintes abreviações: e = entropia, dp = desvio padrão, im = intensidade média, mev4 = mediana do histograma de contraste, qlv4 = primeiro quartil do histograma de contraste, q3v4 = terceiro quartil do histograma de contraste e maxv4 = máximo do histograma de contraste.

Os resultados das validações aplicadas às faces mostraram um bom desempenho, já que foram usadas poucas imagens no conjunto de treinamento e assim a validação *bootstrap* se torna mais adequada à verificação do desempenho dos atributos e também da qualidade do modelo.

Na Tabela 10, nota-se que novamente a intensidade média teve o melhor resultado (o menor erro) usando 6 centros (em média), e a entropia apresentou o pior resultado usando 3 centros (em média). Usando dois atributos combinados, a intensidade média e a mediana de contraste, foi obtido o melhor desempenho dentre todas as combinações possíveis. O modelo usou 24 centros (em média). Por outro lado, a combinação entre o desvio padrão e a mediana do contraste teve o pior desempenho, mesmo usando apenas um centro nas três validações (suavidade máxima).

Combinando três atributos, o melhor resultado foi obtido entre entropia, desvio padrão e intensidade média, com 11 centros (em média). Já o pior resultado foi entre desvio padrão, primeiro quartil e terceiro quartil do histograma de contraste.

Usando a combinação entre quatro atributos, o melhor resultado foi obtido pelos atributos: entropia, desvio padrão, intensidade média e máximo de contraste, usando em média 7 centros.

Novamente o padrão seguido pelos piores resultados, sempre usando poucos centros. Também observando os melhores resultados, nota-se que sempre utilizam um nível de suavização entre o máximo e a média.

#### **4.4. Comparação entre Limiares Automáticos e Humanos**

Nesta seção são apresentados os resultados da análise entre a percepção visual humana aplicada na limiarização de imagens e sua relação com alguns modelos de limiarização automáticos. O objetivo é identificar se há alguma relação entre os limiares humanos e automáticos. Foram comparados os limiares humanos aos limiares automáticos propostos por Kapur [KAPUR et. al., 1985], Brink [BRINK, 1996] e Otsu [OTSU, 1978].



O método de Kapur é baseado na entropia, enquanto o método proposto por Otsu se baseia na variância e Brink por sua vez, propõe um método baseado na entropia em conjunto com a variância. Estes métodos foram escolhidos por utilizarem, como critério de limiarização, a entropia e a variância. Sendo mais específico, o que se quer é identificar alguma correlação entre a entropia/variância e a percepção visual humana aplicada na limiarização de imagens. Esta identificação ocorreu comparando-se os limiares humanos com os obtidos por métodos automáticos.

Com esse objetivo, foi desenvolvido um aplicativo (veja seção 3.5) cujos resultados de sua aplicação são apresentados nas próximas três subseções. Em cada uma destas subseções analisa-se o comportamento de um determinado modelo automático de limiarização de imagens em relação aos limiares humanos.

#### 4.4.1. Comparação entre Limiares: Modelo Kapur e Humanos

A comparação entre os limiares humanos e os limiares automáticos propostos por Kapur [KAPUR et. al., 1985] foi desenvolvida usando o diagrama de dispersão e o coeficiente de correlação entre os limiares humanos e automáticos. O diagrama de dispersão para o conjunto de imagens variadas pode ser observado na Figura 35.

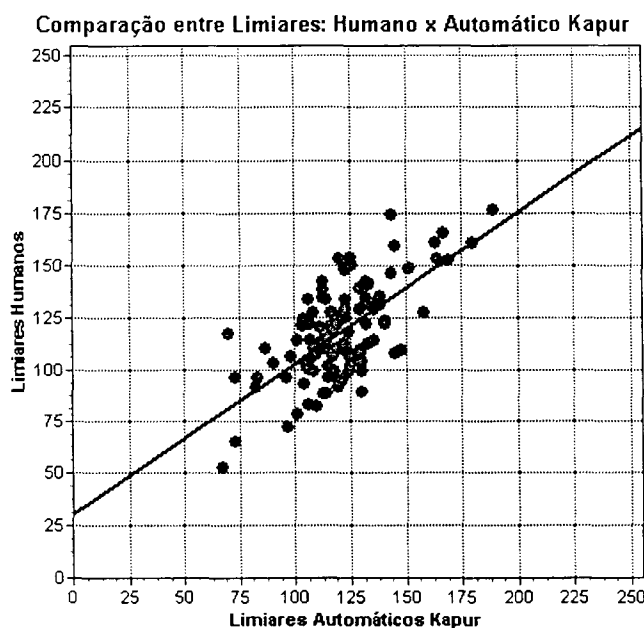


Figura 35 - Diagrama de dispersão entre limiares humanos e automáticos propostos por Kapur, aplicados ao conjunto de imagens variadas.

Como pode ser observado na Figura 35 também está presente no diagrama uma reta de ajuste determinada pelo método dos mínimos quadrados [MOORE, 2000]. O coeficiente de correlação neste conjunto de dados é de 0,6671 o que indica a suspeita de que existe alguma correlação entre os limiares humanos e a entropia (atributo da imagem usado por Kapur).

Uma vez observado que o coeficiente de correlação para o conjunto de imagens variadas foi positivo e maior que 0,5, a comparação também foi desenvolvida para o conjunto de imagens de faces. O diagrama de dispersão entre os limiares humanos e automáticos do conjunto de faces pode ser observado na Figura 36.

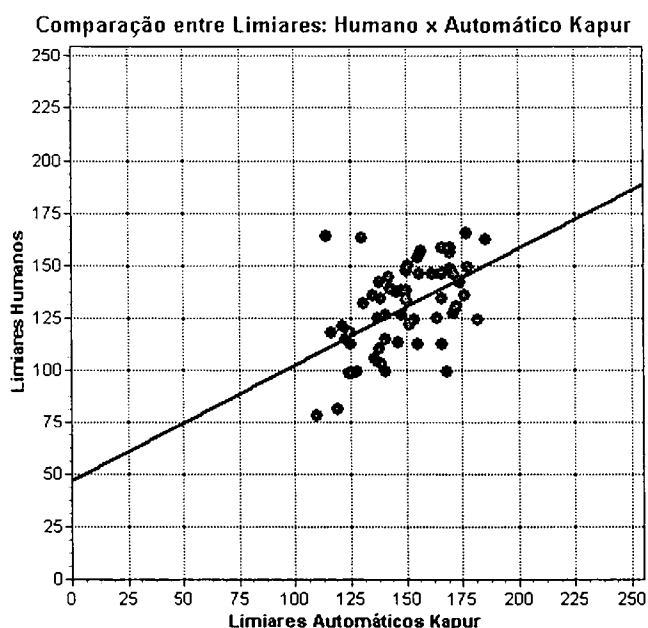


Figura 36 - Diagrama de dispersão entre limiares humanos e automáticos propostos por Kapur, aplicados ao conjunto de imagens de faces.

O coeficiente de correlação deste conjunto de dados é de 0,5066. Este grau de dependência indica também que existe alguma correlação entre os limiares humanos e a entropia. Embora possa se observar a presença de *outliers*, convém questionar a eficiência do método de Kapur em detectar o limiar escolhido por sujeitos humanos. Este coeficiente de correlação se mostra também como um índice de desempenho para métodos automáticos de limiarização.

#### 4.4.2. Comparação entre Limiares: Método Brink e Humanos

Tendo em vista que a comparação entre os limiares humanos e os limiares do modelo Kapur apresentaram alguma correlação (entropia x percepção visual), também foi desenvolvida a mesma análise, mas agora usando os limiares automáticos propostos por Brink [BRINK, 1996], o qual também é baseado na entropia em conjunto com a variância na determinação do limiar, usando para isto uma abordagem *maxmin*.

O diagrama de dispersão confrontando os limiares obtidos pelo método de Brink com os limiares obtidos com o experimento psicofísico para o conjunto de imagens diversas pode ser observado na Figura 37.

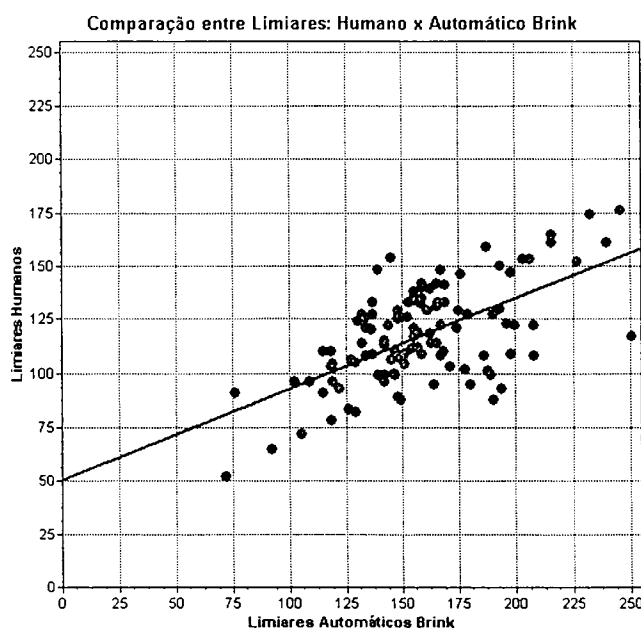


Figura 37 - Diagrama de dispersão entre limiares humanos e automáticos propostos por Brink, aplicados ao conjunto de imagens variadas.

O coeficiente de correlação obtido foi de 0,7104 que mostra também alguma correlação entre os limiares humanos e a entropia e um desempenho similar ao método de Kapur. Contudo com o objetivo de tornar a análise mais consistente, também foi realizada a comparação para o conjunto de imagens de faces, que pode ser observado na Figura 38.

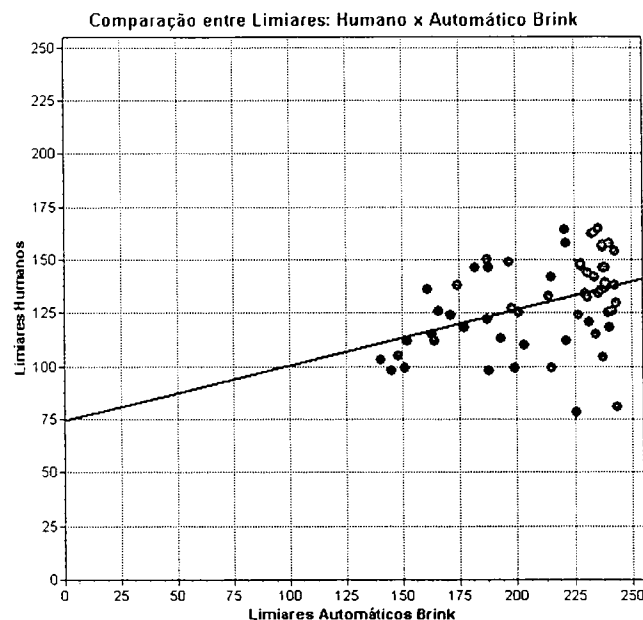


Figura 38 - Diagrama de dispersão entre limiares humanos e automáticos propostos por Brink, aplicados ao conjunto de imagens de faces.

O coeficiente de correlação nesse conjunto de dados foi de 0,512, novamente reforçando a suspeita da existência de uma correlação entre os limiares humanos e a entropia. Esses resultados também indicam uma suspeita de que exista maior correlação entre entropia e limiar em um conjunto de imagens variadas do que em um conjunto homogêneo de imagens, no caso faces.

#### 4.4.3. Comparação entre Limiares: Método Otsu e Humanos

Nesta seção são comparados os resultados obtidos utilizando o método de limiarização de Otsu, que considera a variância da imagem na determinação do limiar, para compará-los aos obtidos com humanos. O que se procura é identificar a existência de uma correlação entre a variância e a percepção visual humana e também avaliar o método de Otsu.

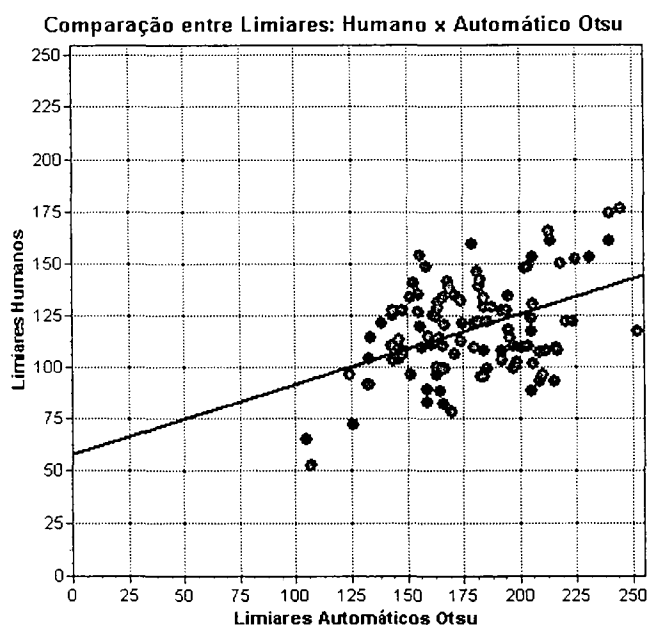


Figura 39 - Diagrama de dispersão entre limiares humanos e automáticos propostos por Otsu, aplicados ao conjunto de imagens variadas.

Foram comparados os limiares humanos e os limiares obtidos com o método automático proposto por Otsu, usando o diagrama de dispersão e o coeficiente de correlação entre os limiares humanos e automáticos. O diagrama de dispersão para o conjunto de imagens variadas é mostrado na Figura 39.

O coeficiente de correlação aplicado ao conjunto de imagens diversas foi de 0,438, indicando uma pequena correlação entre os limiares humanos e a variância. Prosseguindo a investigação, esta comparação foi também realizada usando o conjunto de imagens de faces, com o diagrama de dispersão mostrado na Figura 40.

Neste caso, o coeficiente de correlação foi 0,494, reiterando o baixo desempenho que o método de Otsu apresentou. Este resultado reforça também uma suspeita de que a variância não seja um atributo relevante a um modelo de limiarização baseado na percepção visual.

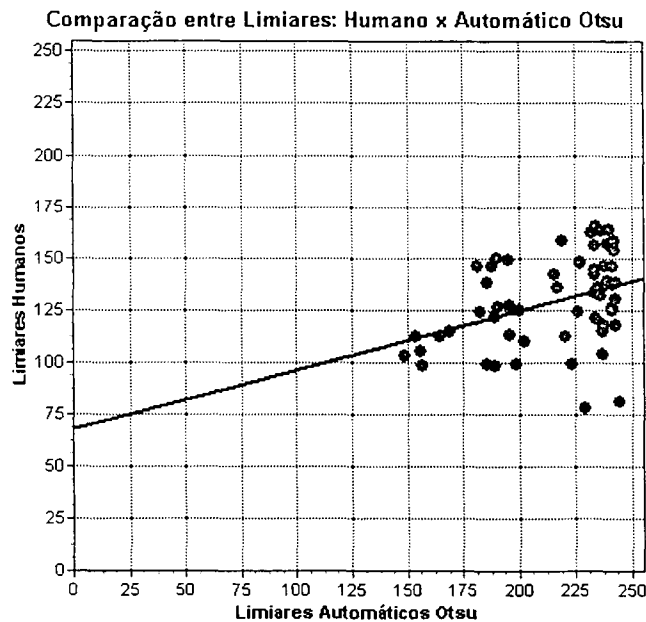


Figura 40 - Diagrama de dispersão entre limiares humanos e automáticos propostos por Otsu, aplicados ao conjunto de imagens de faces.

#### 4.5. Comparação entre Limiares RFBR e Humanos

Nesta seção, os limiares escolhidos por humanos são comparados com aqueles resultantes da aplicação do modelo RFBR. Agora, o objetivo é estimar quais os limites (mínimos e máximos) que podem ser alcançados por modelos baseados nos atributos entropia e variância.

Inicialmente são apresentadas aquelas comparações com os coeficientes de correlação mais alta e na sequência os de mais baixa correlação, usando a entropia. Estes resultados são ilustrados também no diagrama de dispersão e indicados pelo coeficiente de correlação.

Convém definir o que são “os mais altos” e “os mais baixos” coeficientes de correlação, para a generalização com RFBR. Durante a validação *bootstrap*, para cada número de centros são executados 40 testes. As respostas (limiares e erros) do modelo de cada teste são armazenadas e, ao final dos 40 testes, as respostas são classificadas, em ordem crescente, de acordo com os seus erros RMS. Destas respostas classificadas, aquela que apresentou um erro mediano é armazenada em uma tabela.

Os melhores resultados foram aqueles baseados nos limiares medianos mais próximos dos limiares humanos. O diagrama de dispersão contendo os “melhores” limiares do modelo RFBR e os limiares humanos do conjunto de imagens variadas é mostrado na Figura 41.

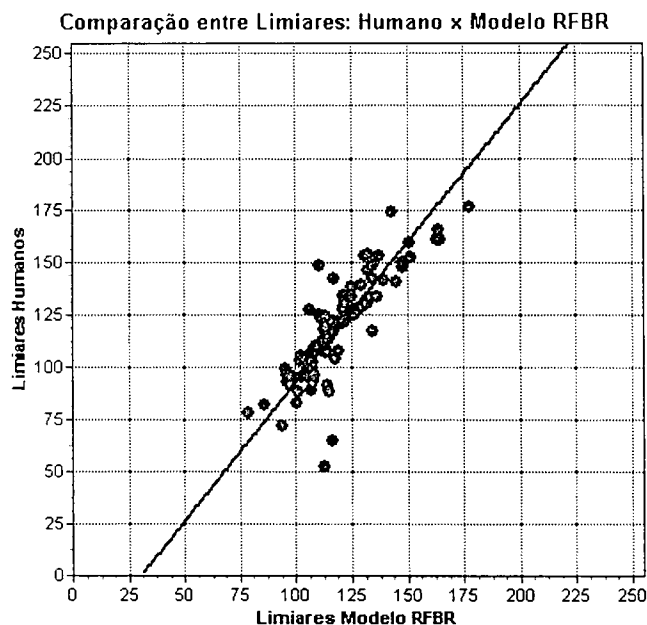


Figura 41 - Diagrama de dispersão entre limiares humanos e melhores limiares automáticos (medianos) do modelo RFBR usando como atributo apenas a entropia, aplicados ao conjunto de imagens variadas.

A Figura 41 mostra a melhor resposta mediana do modelo RFBR indicando um forte grau de correlação (0,84). Este resultado ilustra a flexibilidade de modelos baseados em percepção, já que permite variar a qualidade da resposta. Esta flexibilidade pode ser vista como uma escala de possíveis desempenhos de um método automático de limiarização que utilize um critério baseado em atributos como entropia, variância ou desvio padrão. Convém ressaltar que este resultado pode também ser visto como um limite superior para este tipo de método. Este resultado também reforça a suspeita da existência de mecanismos neurais que podem ser modelados para estabelecer limiares em imagens.

Na escala de possíveis desempenhos também há lugar para os piores resultados. Neste caso, o pior resultado da mediana dos erros RMS, isto é, um coeficiente de correlação de 0,1544 pode ser visualizado na Figura 42.

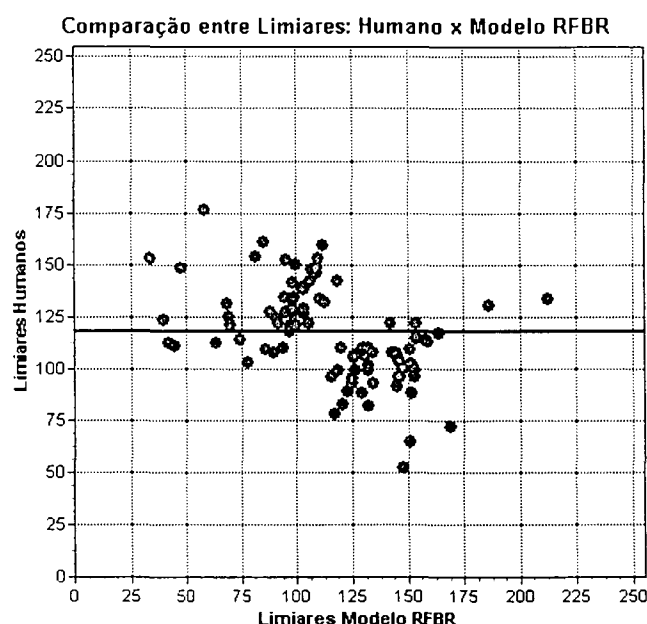


Figura 42 - Diagrama de dispersão entre limiares humanos e piores limiares automáticos (medianos) do modelo RFBR usando como atributo apenas a entropia, aplicados ao conjunto de imagens variadas.

Também foram realizadas comparações entre os desempenhos obtidos pelo atributo variância. Dessa forma, é apresentado o diagrama de dispersão entre os melhores limiares medianos do modelo RFBR (usando a variância) e os limiares humanos, o qual pode ser visto na Figura 43.

Pode-se observar na Figura 43, assim como no diagrama do atributo entropia, que existe um alto grau de correlação entre os limiares humanos e a variância, um coeficiente de correlação de 0,9167. Este resultado também pode servir como uma referência superior no desempenho de métodos automáticos de limiarização baseados no atributo variância. Da mesma forma, fazendo o papel de fundo de escala está um resultado com baixo grau de correlação (Figura 44), ou seja, o pior resultado da mediana do erro RMS, um coeficiente de correlação de 0,0937.



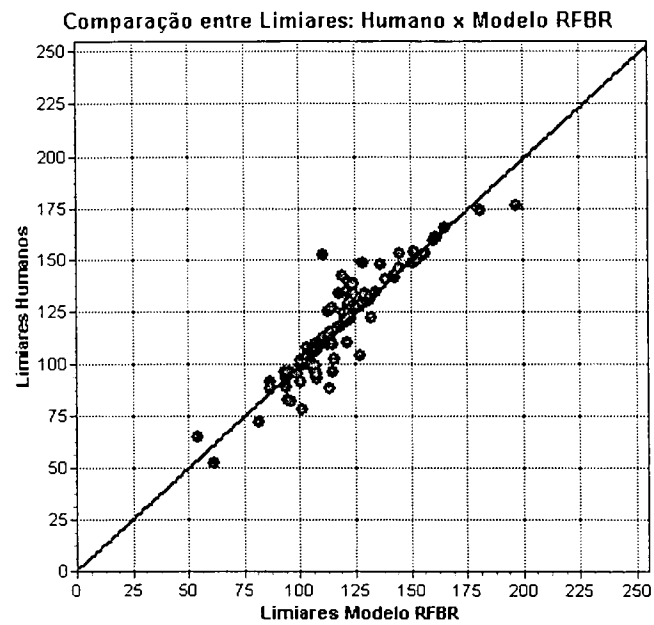


Figura 43 - Diagrama de dispersão entre limiares humanos e melhores limiares automáticos (medianos) do modelo RFBR usando como atributo apenas a variância, aplicados ao conjunto de imagens variadas.

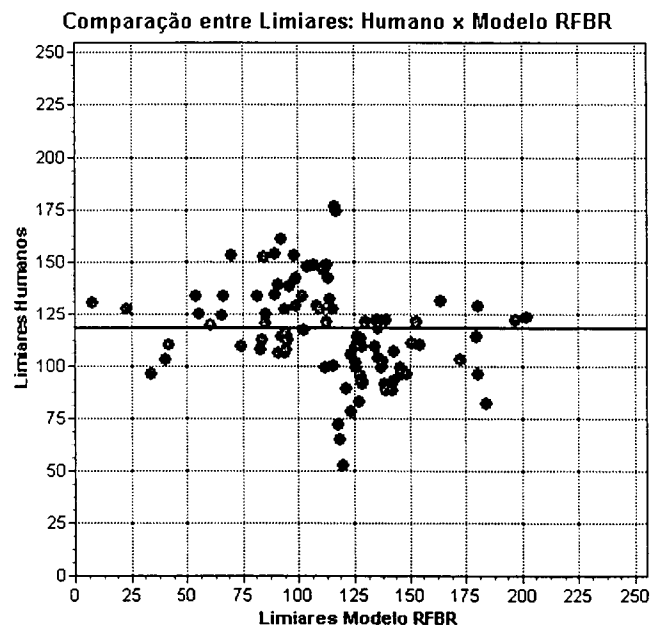


Figura 44 - Diagrama de dispersão entre limiares humanos e piores limiares automáticos (medianos) do modelo RFBR usando como atributo apenas a variância, aplicados ao conjunto de imagens variadas.

## 5 CONCLUSÃO

Este trabalho apresentou uma abordagem para a investigação de métodos de limiarização de imagens digitais baseados na percepção visual humana. Esta abordagem generaliza a resposta de sujeitos humanos à limiarização manual obtida em um experimento psicofísico. Esta generalização foi realizada de duas maneiras: uma com a interpolação com thin-plate spline (TPS), mais rígida; e outra com uma generalização usando uma rede de funções de base radial (RFBR). As respostas dos modelos obtidos com a RFBR foram validadas com três métodos: (i) deixa-um-fora, que oferece uma validação mais pessimista; (ii) a cruzada generalizada (conhecida também como k-dobras); (iii) a bootstrap, estas duas últimas mais realistas. Estas validações foram aplicadas também em dados gerados por funções de controle com comportamento linear, exponencial, logarítmico, polinomial e aleatório. Foi observado o comportamento das validações, analisando os erros RMS, para diversas execuções de treinamento/teste, em função do número de centros utilizados na generalização com RFBR. Observou-se que os erros RMS decrescem quando aplicados os modelos sobre as funções de controle, exceto a função aleatória que se observa um crescimento dos erros RMS. Isto acontece porque a exatidão do modelo exige um maior número de funções para aproximar, por exemplo, uma reta. Esta propriedade das RFBR permitiria construir, em um trabalho futuro, um ajuste de certeza para os modelos.

Foram analisados também os resultados obtidos com técnicas automáticas de limiarização e os limiares coletados no experimento psicofísico com sujeitos humanos. Os resultados mostraram alguma correlação entre os limiares obtidos por estas técnicas e as do experimento, contudo estes resultados não se mostraram tão similares às respostas perceptivas. Algumas publicações chegam a preconizar avaliações visuais deste tipo de técnica [BEGHDADI et. al., 1995], [BRINK, 1996], [BRINK & PENDOCK, 1996], [BRINK, 1992], [YANOWITZ & BRUCKSTEIN, 1989] e [YIN, 2002].

Portanto, a abordagem de modelos usando RFBRs se mostrou, além de um método de limiarização baseado na percepção visual humana, uma referência para

avaliação de métodos automáticos de limiarização, pois é capaz de confrontá-los quantitativamente com as respostas perceptivas. Esta capacidade de definir quantitativamente a qualidade de um método de limiarização permitirá também desenvolver, no futuro, um seletor de características que, acoplado a uma estratégia de busca (por exemplo, uma busca evolucionária) poderá definir quais atributos são mais relevantes, desde que seja estabelecido um grau de incerteza aceitável pelo modelo.

Este trabalho contribui também com um experimento psicofísico capaz de colher limiares atribuídos por humanos, orientado ao uso em ambientes web, permitindo que grandes audiências participem dos experimentos. Em uma próxima versão deste ambiente, recomenda-se a implementação do método dos limites (veja seção 2.5) nos experimentos, com o intuito de garantir que o sujeito não sofra influência dos aspectos visuais da GUI. Recomenda-se também que os próximos experimentos possam ser realizados sob controle mais rígido das condições ambientais tais como iluminação, posição do monitor, ajuste de contraste e brilho do monitor, entre outras variáveis.

## 6 REFERÊNCIAS

- ABUTALEB, A.H. Automatic Thresholding of Gray-Level Pictures Using Two-Dimensional Entropy. *Computer Vision, Graphics, and Image Processing*, (47), p22-32, 1989.
- AHUJA, N.; ROSENFELD, A. A Note on the Use of Second-Order Gray-Level Statistics for Threshold Selection, *IEEE Trans. Systems Man Cybernet.*, (8), p895-899, 1978.
- BEGHDADI, A.; NÉGRATE, A.L.; DE LESEGNO, P.V. Entropic Thresholding Using a Block Source Model. *CVGIP: Graphical Models and Image Processing*, (57), p197-205, 1995.
- BISHOP, C.M. *Neural Networks for Pattern Recognition*. Oxford University Press, 482pp., 1995.
- BOLTZMANN, L. *The Second law of thermodynamics (1886)*. Reprinted In: MCGUINNESS, B. (Ed.) Boltzmann, Theoretical Physics and Philosophical Problems. New York: D. Reidel, 1974.
- BRINK, A.D. Using spatial information as an aid to maximum entropy image threshold selection. *Pattern Recognition Letters*, (17), p29-36, 1996.
- BRINK, A.D. Thresholding of Digital Images Using Two-Dimensional Entropies. *Pattern Recognition*, (25), p803-808, 1992.
- BRINK, A.D.; PENDOCK, N.E. Minimum Cross-Entropy Threshold Selection. *Pattern Recognition*, (29), p179-188, 1996.
- BROOMHEAD, D.S.; LOWE, D. Multivariable Functional Interpolation and Adaptive Networks. *Complex Systems*, (2), p321-355, 1988.
- CASTLEMAN, K. *Digital Image Processing*. Prentice Hall, 667pp., 1996.
- CHENG, H.D.; CHEN, Y.H. Fuzzy partition of two-dimensional histogram and its application to thresholding. *Pattern Recognition*, (32), p825-843, 1999.
- CHUN, D.N.; YANG, H.S. Robust image segmentation using genetic algorithm with a fuzzy measure. *Pattern Recognition*, (29), p1195-1211, 1996.

- CONSULARO, L.A. *Técnicas de Mineração de Dados para Análise de Imagens*. Tese (Doutorado) - Instituto de Física de São Carlos, 236pp., 2000.
- DOYLE, W. Operation Useful for Similarity-Invariant Pattern Recognition. *Journal Association Computer Machine*. (9), p259-267, 1962.
- DUDA, R.O.; HART, P.E.; STORK, D.G. *Pattern Classification and Scene Analysis. Part I: Pattern Classification*. John Wiley and Sons, 482pp., 1998.
- EFRON, B.; TIBSHIRANI, R. *An introduction to the bootstrap*. Chapman & Hall, 1993.
- FILHO, J.M. *Instalações Elétricas Industriais*. LTC - Livros Técnicos e Científicos, 753pp., 2001.
- FLEURY, M.; HAYAT, L.; CLARCK, A.F. Parallel Entropic Auto-Thresholding. *Image and Vision Computing*, (14), p247-263, 1996.
- FU, K.S.; MUI, J.K. A Survey on Image Segmentation. *Pattern Recognition*, (13), p3-16, 1981.
- GONZALEZ, R.C.; WOODS, R.E. *Processamento de Imagens Digitais*. Edgard Blücher, 509pp., 2000.
- GORDON, I.E. *Theories of Visual Perception*. John Wiley and Sons, 273pp., 1995.
- GHOSH, J.; NAG, A. An Overview of Radial Basis Function Networks. In: HOWLETT, R.J. & JAIN, L.C. (Eds.) **Radial Basis Function Neural Network Theory and Applications**, Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Heidelberg, Alemanha, p1-36, 2001.
- HARALICK, R.M.; SHANMUGAM, K.; DINSTEN, I. Texture Features for Image Classification, *IEEE Trans. Systems Man Cybernet.*, (3), p610-621, 1973.
- HAYKIN, S. *Redes Neurais: Princípios e Prática*. Bookman, 898pp., 2001.
- JAWAHAR, C.V.; BISWAS, P.K.; RAY, A.K. Analysis of fuzzy thresholding schemes. *Pattern Recognition*, (33), p1339-1349, 2000.
- JOHANNSEN, G.; BILLE, J. A Threshold Selection Method Using Information Measures. *Proceedings 6<sup>th</sup> Int. Conf. Pattern Recognition*, Munich, Germany, p140-143, 1982.

- KAEHLING, L.P.; LITTMAN, M.L.; MOORE, A.W. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, (4), p237-285, 1996.
- KAPUR, J.N.; SAHOO, P.K.; WONG, K.C. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, (29), p273-285, 1985.
- KIRBY, R.L.; ROSENFELD, A. A Note on the Use of (Gray Level, Local Average Gray Level) Space as an Aid in Thresholding Selection, *IEEE Trans. Systems Man Cybernet.*, (9), p860-864, 1979.
- KOHAVER, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *IJCAI - International Joint Conference on Artificial Intelligence*, 1995.
- LEE, S.U.; CHUNG, S.Y.; PARK, R.H. A Comparative Performance Study of Several Global Thresholding Techniques for Segmentation. *Computer Vision, Graphics, and Image Processing*, (52), p171-190, 1990.
- LEVINE, D. M.; NAZIF, A. M. Dynamic Measurement of Computer Generated Image Segmentations. *IEEE Trans. Pattern Anal. Mach. Intell. PAMI*, (7), p144, 164, 1985.
- MASON, D.; LAUDER, I.J.; RUTORITZ, D.; SPOWART, G. Measurement of C-Bands in Human Chromosomes, *Comput. Biol. Med.*, (5), p179-201, 1975.
- MEES, A.I.; JACKSON, M.F.; CHUA, L.O. Device Modeling by Radial Basis Functions. *IEEE Transactions on Circuits and Systems - I: Fundamental Theory and Applications*, (39)1, 19-27, 1992.
- MOORE, D. *A estatística básica e sua prática*. LTC - Livros Técnicos e Científicos, 482pp., 2000.
- ORR, M.J.L. *Introduction to Radial Basis Function Networks*. Relatório Técnico, University of Edinburgh, 67pp. 1996. [<http://anc.ed.ac.uk/~mjo/rbf.html>]
- OTSU, N. A threshold selection method from gray-level histogram. *IEEE Transactions on Systems, Man, and Cybernetics*, p62-66, 1978.
- PAL, N.R.; PAL, S.K. A Review on Image Segmentation Techniques. *Pattern Recognition*, (26), p1277-1294, 1993.

- PEDRYCZ, W. *Computational intelligence : an introduction*. CRC Press, 284pp., 1998.
- POWELL, M.J.D. *Radial basis functions for multivariate interpolation: A review*, Relatório Técnico 1985/NA12. Department of Applied Mathematics and Theoretical Physics, Cambridge University, 1985.
- PREWIT, J.M.S.; MENDELSON, M.L. The Analysis of Cell Images, *Ann. New York Acad. Sci.* (128), p1035-1053, New York Acad. Sci., New York, 1966.
- PUN, T. A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram. *Signal Processing*, (2), p223-237, 1980.
- ROSENFELD, A.; DE LA TORRE, P. Histogram Concavity Analysis as an Aid in Threshold Selection. *IEEE Transactions on Systems, Man, and Cybernetics*. (13), p231-235, 1983.
- SAHOO, P.K.; SOLTANI, S.; WONG, A.K.C. A Survey of Thresholding Techniques. *Computer Vision, Graphics, and Image Processing*, (41), p233-260, 1988.
- SANDWELL, D.T. Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data. *Geophysical Research Letters*, (2), 139-142, 1987.
- SHANNON, C.E. A Mathematical Theory of Communication. *Bell System Technical Journal*, (27)3, p379-423, 1948.
- STEVENS, C.F. *The Six Core Theories of Modern Physics*. MIT Press, 221pp., 1998.
- STEVENS, S.S. *Psychophysics: introduction to its perceptual, neural and social prospects*. Transaction, 329pp., 1986.
- THEODORIDIS, S., KONSTANTINOS, K., *Pattern recognition*, 1st edition. Academic press, 625p., 1999.
- WATKINS, C.J.C.H.; DAYAN, P. Q-Learning. *Machine Learning*, (3), p279-292, 1992.
- WESKA, J. S. A Survey of Thresholding Selection Techniques. *Computer Graphics and Image Processing*, (5), p382-399, 1978.
- WU, A.Y.; ROSENFELD, A. Threshold Selection Using Quadtree, *IEEE Trans. Pattern Anal. Mach. Intell.*, (4), p90-94, 1982.

- YANOWITZ, S.D.; BRUCKSTEIN, A.M. A New Method for Image Segmentation. *CVGIP*, (46), p82-95, 1989.
- YIN, P.Y. A Fast Scheme for Optimal Thresholding Using Genetic Algorithms. *Sinal Processing*, (72), p85-95, 1999.
- YIN, P.Y. Maximum Entropy-Based Optimal Threshold Selection Using Deterministic Reinforcement Learning With Controlled Randomization. *Sinal Processing*, (82), p993-1006, 2002.
- ZHANG, Y. J. A Survey on Evaluation Methods for Image Segmentation. *Pattern Recognition*, (31), p1335-1346, 1996.



## 7 APÊNDICE

### 7.1. Imagens Variadas

Nesta seção é disponibilizada uma tabela (Tabela 11) contendo o conjunto de imagens variadas, sendo cada uma das imagens identificada pelo nome e acompanhada pelos respectivos atributos e limiares, os quais foram usados durante a realização deste trabalho.

nome da imagem	entropia	desvio padrão	intensidade média	quartil1v4	medianav4	quartil3v4	maximov4	Limiar Humano	limiar kapur	limiar brink	limiar Otsu
paisagem4	5,844	23,386	149,624	3	5	8	78	135	139	159	155
quarto	5,887	36,093	187,016	1	1	2	48	161	163	216	214
casa2	6,054	86,028	174,222	0	2	7	115	176	189	246	245
quarto2	6,109	34,222	187,937	0	1	2	61	165	167	216	213
paisagem3	6,337	25,967	148,008	2	3	7	76	140	133	159	153
5112	6,404	57,249	185,98	1	2	5	86	152	169	227	225
casa3	6,446	90,796	159,737	0	3	10	136	161	180	240	240
yugo	6,491	92,276	141,807	0	4	8	114	117	70	251	252
paisagem1	6,494	31,269	165,675	2	4	8	78	159	145	187	179
paisagem5	6,549	28,49	131,1	2	3	6	92	104	130	151	147
casa1	6,557	79,619	148,11	1	2	9	95	174	144	233	240
picture4	6,634	41,742	113,9	2	2	6	102	125	123	133	144
camera	6,77	62,341	118,724	2	3	10	136	114	136	163	165
picture43	6,822	47,794	69,6768	2	3	6	90	91	82	76	132
boat2	6,869	51,593	135,751	1	4	12	78	141	134	169	168
boat	6,879	45,7	129,334	2	6	12	107	133	123	153	152
casa4	6,893	74,007	150,995	2	7	13	120	153	164	203	206
4101	6,895	39,751	58,3576	2	3	6	86	52	67	72	107
jvz00339	6,899	72,157	109,273	2	5	16	163	127	117	179	194
brezhnev	6,911	72,842	115,009	3	10	21	217	117	114	155	205
bullsceltics	6,926	71,465	86,4272	4	8	23	232	96	73	102	210
picture46	6,935	41,519	114,189	2	4	8	114	114	107	132	134
paisagem2	6,946	50,673	137,36	2	4	8	86	131	139	165	164
6101	6,979	54,236	87,2685	1	2	4	62	104	106	119	133
borboleta2	6,985	43,616	102,353	1	2	9	116	91	119	115	133
picture32	7,023	78,841	108,469	2	7	21	156	148	152	167	204
picture40	7,029	39,62	75,2542	2	4	9	109	65	73	92	105
picture12	7,037	73,677	131,052	2	4	10	129	123	141	196	205
jvz00328	7,063	65,979	102,072	2	6	15	138	129	123	148	184
picture37	7,082	50,815	139,17	2	4	9	152	121	104	174	175
picture33	7,099	55,872	136,301	2	2	5	103	95	121	180	184

picture34	7,101	49,372	113,44	2	4	10	118	121	103	134	139
picture20	7,121	56,988	93,7524	2	6	15	116	96	96	109	163
mandrill	7,131	38,829	132,764	6	12	20	98	119	118	157	156
manhunt	7,133	46,22	106,875	4	6	12	119	110	119	118	143
picture14	7,143	71,269	135,838	2	4	8	102	108	125	208	217
jets2	7,157	77,787	88,0363	3	6	11	193	93	104	122	209
picture24	7,168	57,021	114,261	5	11	21	165	82	110	129	166
picture10	7,18	55,804	131,424	3	6	12	166	125	121	148	162
zelda	7,191	40,288	90,7965	2	4	6	74	96	83	119	124
picture13	7,199	44,607	86,248	4	9	17	158	72	97	105	126
picture26	7,227	41,683	144,459	6	11	19	182	133	138	169	166
picture1	7,233	59,89	128,295	3	6	12	228	109	122	159	180
bailout	7,234	78,712	136,995	1	3	8	132	147	123	198	202
tulipa1	7,245	49,874	126,798	2	4	7	87	154	125	145	156
picture2	7,246	70,225	129,633	2	3	7	165	88	113	190	205
picture5	7,249	57,277	122,508	2	6	16	236	120	111	136	167
picture16	7,255	49,423	121,867	5	13	24	188	109	129	137	157
picture11	7,276	49,417	110,594	6	10	18	141	127	108	137	144
picture31	7,284	66,765	128,198	2	6	13	106	109	115	151	216
picture23	7,289	61,554	123,376	2	5	11	155	106	117	145	171
goldhill	7,301	49,071	112,212	4	8	14	95	105	125	129	147
picture35	7,309	52,183	111,453	3	6	14	96	127	116	132	148
barb	7,317	45,283	112,071	2	6	16	74	113	112	142	147
lena	7,328	47,378	123,677	2	4	8	82	126	122	152	155
picture17	7,329	46,54	110,066	7	12	20	148	108	117	134	148
montanha	7,332	74,099	142,035	4	12	22	168	153	120	206	231
picture29	7,34	50,753	146,666	4	9	16	164	146	144	176	181
profile9	7,342	64,967	110,967	1	2	8	128	122	107	144	186
picture38	7,355	61,651	118,145	2	6	15	92	99	118	139	186
elaine	7,366	45,586	135,985	3	4	7	72	129	136	161	164
picture28	7,372	54,592	128,938	4	10	19	135	134	132	155	172
altschul	7,377	72,906	85,272	3	10	25	153	110	87	115	197
picture25	7,386	48,945	122,613	6	14	24	162	89	130	148	159
picture27	7,391	48,409	132,917	3	6	15	199	112	133	157	163
picture18	7,401	61,018	97,5539	3	8	16	92	78	101	118	170
picture19	7,402	52,289	112,745	6	12	20	164	96	115	142	152
homeless1	7,402	62,846	148,565	1	2	5	96	109	148	198	201
chile	7,403	53,866	156,203	3	5	11	152	127	158	190	192
cancer	7,407	78,386	99,1244	2	6	13	145	103	105	171	192
huac	7,415	64,897	120,112	2	4	12	162	118	124	162	195
picture30	7,417	54,324	112,982	7	16	29	164	99	124	142	166
pimentoes	7,423	52,715	115,978	2	4	8	79	111	114	154	161
picture42	7,43	55,869	102,028	2	6	14	106	83	106	126	159
rdp	7,43	58,584	89,2332	2	6	13	120	103	90	118	144
picture41	7,435	57,644	96,9034	2	7	14	99	124	104	130	163
johndean	7,437	61,509	87,0543	2	6	16	192	106	98	127	148
picture22	7,44	62,109	132,406	2	6	15	117	108	145	167	192
bridge	7,452	53,109	113,855	7	12	19	94	148	123	139	159
picture15	7,467	53,617	124,427	3	6	14	114	132	132	158	174
tulipa2	7,468	56,69	107,135	3	5	8	85	133	106	137	152

afghanistan	7,473	89,631	125,696	3	6	20	172	122	141	200	223
picture3	7,479	71,852	134,448	2	6	14	104	93	120	194	215
annan	7,488	76,838	121,139	2	6	14	119	101	115	188	206
coruja	7,52	70,883	117,929	3	8	16	103	107	109	149	209
nichols	7,524	75,598	112,317	3	6	15	101	142	132	165	183
profile5	7,533	72,943	112,408	2	4	10	103	114	101	165	196
asiaecon	7,537	62,104	103,829	3	7	18	136	115	115	142	160
objectfilmreview	7,552	86,018	130,932	2	5	12	106	122	132	208	220
picture21	7,57	62,714	140,393	6	12	21	178	99	130	189	197
hat	7,577	65,335	119,418	4	8	16	117	121	116	155	180
picture39	7,587	56,175	115,353	6	11	21	156	110	122	147	166
docworld	7,596	81,429	120,323	2	6	18	145	108	129	186	212
filmmakers	7,604	71,809	116,301	3	8	26	168	139	129	169	182
stern	7,605	64,19	109,264	2	6	13	108	138	113	155	169
rusiaecon	7,606	74,039	111,612	2	5	14	152	110	108	168	204
brief	7,606	57,362	109,308	3	9	21	121	99	108	147	167
vpluto	7,611	61,676	114,807	2	5	12	105	112	122	155	174
picture6	7,639	57,529	113,927	4	7	15	104	88	114	149	165
horsemeat	7,643	67,863	115,106	5	9	15	128	108	110	148	184
macau	7,644	75,181	105,097	3	7	14	114	134	113	156	195
curriel	7,659	78,668	114,593	4	7	13	142	102	125	178	199
airforce	7,66	61,937	105,847	4	10	21	124	100	106	146	163
Ray	7,662	70,147	113,166	3	6	14	112	133	114	166	184
mafiareview	7,665	81,861	125,967	2	6	13	169	150	126	193	218
oscars	7,684	80,065	129,453	4	9	22	118	130	137	193	206
philjackson	7,733	65,677	117,406	3	6	12	110	122	119	167	182
cthartford	7,733	70,803	115,893	4	10	20	162	142	113	159	183
gloup	7,75	68,884	128,11	7	17	34	161	129	129	175	188
capitol	7,794	68,19	120,126	3	8	18	110	95	119	164	183

Tabela 11 - Conjunto de imagens de faces, seus atributos e limiares.

## 7.2. Imagens de Faces

Nesta seção é disponibilizada uma tabela (Tabela 12) contendo o conjunto de imagens de faces, sendo cada uma das imagens identificada pelo nome e acompanhada pelos respectivos atributos e limiares, os quais foram usados durante a realização deste trabalho.

nome da imagem	entropia	desvio padrão	Intensidade média	quartil1v4	medianav4	quartil3v4	maximov4	Limiar Humano	limiar kapur	limiar brink	limiar Otsu
49	6,13	67,244	190,86	1	4	12	78	130	173	244	243
29	6,294	63,31	140,74	2	3	9	80	146	166	182	181
31	6,366	66,422	192,27	1	2	6	87	124	182	227	226
43	6,394	59,247	200,1	1	3	8	116	162	186	233	232
21	6,412	65,423	191,99	1	3	9	56	154	155	243	242
35	6,453	54,241	173,18	2	4	10	86	146	162	188	188
7	6,467	70,155	183,66	1	2	8	81	126	141	242	241
25	6,483	59,772	167,92	2	5	13	108	149	178	197	195
23	6,502	92,534	151,61	1	3	8	101	158	166	241	242
9	6,532	93,353	149,37	1	3	9	114	137	146	239	241
20	6,564	64,201	178,65	1	4	11	88	139	143	239	239
39	6,6	98,809	147,97	1	2	8	134	146	171	238	241
12	6,616	50,324	162,78	1	4	11	80	110	138	203	202
6	6,622	88,674	153,33	1	4	12	106	81	119	244	244
53	6,692	70,521	174,37	2	6	16	103	138	148	243	243
27	6,714	92,852	129,68	1	2	10	96	99	168	215	223
16	6,75	62,254	191,63	1	3	8	95	136	176	237	235
45	6,759	66,61	162,9	2	4	10	98	121	121	232	234
24	6,773	65,4	182,97	1	4	11	116	104	137	238	237
54	6,785	65,195	162,41	2	8	17	103	142	138	215	215
19	6,794	60,583	165,23	1	2	8	81	99	141	199	199
14	6,819	84,844	149,7	1	4	10	104	118	125	241	243
33	6,83	51,371	122,66	2	5	12	102	103	139	140	149
5	6,833	80,155	152,88	2	4	11	109	115	123	235	237
10	6,854	61,731	177,56	2	4	10	92	147	150	228	227
52	6,866	63,837	183,71	2	5	12	78	146	156	239	238
55	6,888	73,987	166,23	2	6	17	133	125	137	240	241
11	6,906	84,865	168,44	1	2	7	89	134	166	236	235
47	6,921	63,958	176,66	2	4	10	113	148	170	228	227
40	6,937	69,313	180,56	1	6	12	83	156	170	238	233
18	6,938	58,783	139,33	1	3	9	72	124	154	171	183
4	6,94	81,267	161,62	2	5	13	88	157	157	238	239

50	6,947	53,621	162,29	2	6	15	116	122	152	187	189
15	6,951	57,02	129,03	3	6	12	90	112	125	164	165
37	6,955	59,533	138,72	2	3	6	63	115	141	163	169
56	6,958	59,677	138,44	2	4	8	91	138	150	174	186
26	6,976	52,46	124,06	2	5	11	84	105	136	148	156
48	6,994	73,669	179,34	2	6	16	143	142	174	234	233
1	6,994	50,763	132,02	2	4	11	58	112	155	152	154
58	7,038	86,428	144,56	1	4	10	100	132	131	231	236
2	7,047	66,383	181,59	2	8	24	154	165	177	236	234
38	7,054	72,727	137,23	2	5	11	58	118	116	177	238
34	7,06	57,122	181,39	1	5	11	74	158	170	222	219
46	7,067	79,622	114,47	2	4	11	99	136	135	161	217
59	7,071	52,729	170,91	4	6	10	115	127	171	198	196
8	7,078	84,901	142,61	2	4	14	93	133	150	214	233
57	7,095	56,539	170,04	3	6	12	102	125	164	201	200
32	7,119	86,538	150,38	2	7	22	106	163	130	234	236
36	7,124	58,552	156,64	2	4	9,25	91	98	124	188	189
3	7,124	62,457	177,65	2	4	10	103	112	166	222	220
13	7,154	83,551	138,73	2	4	10	110	78	110	226	229
17	7,158	61,227	102,12	2	4	10	74	98	126	145	157
51	7,168	69,681	164,95	2	5	13	73	134	139	230	234
42	7,175	86,297	155,29	2	5	13,25	107	144	142	231	233
41	7,201	71,851	113,89	2	4	9	61	126	148	166	191
44	7,211	86,947	147,76	1	5	13	80	164	114	221	240
30	7,217	64,852	119,56	2	6	16	98	99	128	151	186
22	7,294	58,691	144,33	2	4	9	92	150	151	187	190
28	7,368	65,467	150,03	3	6	13	105	113	147	193	196

Tabela 12 - Conjunto de imagens de faces, seus atributos e limiares.